

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:36:07 ; Search time 33.27 Seconds  
(without alignments)  
35.790 Million cell updates/sec

Title: US-09-489-760-5  
Perfect score: 42  
Sequence: 1 VLRLDLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL16.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_plant.\*
- 10: sp\_please.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42	100.0	1165	4 Q92619	Q92619 homo sapien
2	38	90.5	4150	2 Q9KIV3	Q9KIV3 streptomyc
3	36	85.7	174	5 Q94256	Q94256 caenorhabdi
4	35	83.3	616	1 Q27025	Q27025 methanobact
5	34	81.0	238	14 Q9WHX7	Q9WHX7 influenza a
6	34	81.0	239	14 Q9WHY9	Q9WHY9 influenza a
7	34	81.0	239	14 Q9WHY8	Q9WHY8 influenza a
8	34	81.0	239	14 Q9WHY7	Q9WHY7 influenza a
9	34	81.0	239	14 Q9WHY6	Q9WHY6 influenza a
10	34	81.0	239	14 Q9WHY5	Q9WHY5 influenza a
11	34	81.0	239	14 Q9WHY4	Q9WHY4 influenza a
12	34	81.0	239	14 Q9WHY3	Q9WHY3 influenza a
13	34	81.0	239	14 Q9WHY1	Q9WHY1 influenza a
14	34	81.0	239	14 Q9WHX9	Q9WHX9 influenza a
15	34	81.0	250	14 Q9WHY2	Q9WHY2 influenza a
16	34	81.0	252	14 Q93188	Q93188 influenza a
17	34	81.0	389	2 Q9XTD7	Q9XTD7 caenorhabdi
18	34	81.0	486	2 Q9PPU0	Q9PPU0 ureaplasma
19	34	81.0	791	5 Q9NG78	Q9NG78 caenorhabdi

20	34	81.0	3816	2 Q9KIV3	Q9KIV3 streptomyc
21	33	78.6	300	2 Q9KFG3	Q9KFG3 bacillus ha
22	33	78.6	324	10 P93619	P93619 vigna ungui
23	33	78.6	377	8 O09424	O09424 chrysosplen
24	33	78.6	392	8 Q36681	Q36681 quintinia v
25	33	78.6	396	8 O09418	O09418 chrysosplen
26	33	78.6	396	8 O09426	O09426 chrysosplen
27	33	78.6	432	8 O78605	O78605 isotria ver
28	33	78.6	435	8 O09416	O09416 chrysosplen
29	33	78.6	437	8 O09422	O09422 chrysosplen
30	33	78.6	440	8 Q33148	Q33148 saxifraga p
31	33	78.6	441	8 Q31822	Q31822 antrophum
32	33	78.6	465	8 Q31907	Q31907 bolandra or
33	33	78.6	465	8 Q32207	Q32207 elmera race
34	33	78.6	465	8 Q32484	Q32484 jepsonia pa
35	33	78.6	465	8 Q33136	Q33136 sullivantia
36	33	78.6	465	8 Q33509	Q33509 hydrangea m
37	33	78.6	466	8 O20245	O20245 achlys trip
38	33	78.6	467	8 Q9TLN7	Q9TLN7 asteropyrum
39	33	78.6	469	8 Q9GDP7	Q9GDP7 quintinia v
40	33	78.6	475	8 Q9MTF0	Q9MTF0 tofieldia p
41	33	78.6	833	10 Q9LV01	Q9LV01 arabidopsis
42	32	76.2	237	2 O06978	O06978 bacillus su
43	32	76.2	274	5 Q9KL57	Q9KL57 drosophila
44	32	76.2	389	2 O86581	O86581 streptomyc
45	32	76.2	452	3 O43088	O43088 schizosacch

## ALIGNMENTS

## RESULT 1

Q92619 ID Q92619 PRELIMINARY; PRT: 1165 AA.  
AC Q92619;  
DT 01-FEB-1997 (TEMBLrel. 02, Created)  
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)  
DE 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
DE MYELOBLAST KIAA0223 (FRAGMENT).  
GN KIAA0223.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BONE MARROW;

RX MEDLINE=97191544; PubMed=9039502; Ohira M., Kawarabayasi Y., Ohara O.,

RA Nagase T., Seki N., Ishikawa K., Nomura N.;

RA Tanaka A., Kotani H., Miyajima N., Nomura N.;

RT "Prediction of the coding sequences of unidentified human genes. VI.

RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by

RT analysis of cDNA clones from cell line KG-1 and brain.;"

RL DNA Res. 3:321-329(1996).

RN [2]

RP SEQUENCE OF 170-1165 FROM N.A.

RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,

RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,

RA Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,

RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,

RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,

RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,

RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,

RA Carrano A.V.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; D66976; BAA13212.1; -;

DR EMBL; AC004151; AAC03237.1; -;

DR HSSP; Q07960; IRGP.

DR InterPro; IPR000198; -;

DR InterPro; IPR002219; -;

DR Pfam; PF001130; DAG\_PE-bind; 1.

DR Pfam; PF00620; RhoGAP; 1.

DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; UNKNOWN\_1.

```

DR PROSITE; PS0081; DAG_PE_BIND_DOM_2; 1.
DR SMART; SM00324; RhogAP; 1.
FT NON_TER 1
SQ SEQUENCE 1165 AA; 127344 MW; 92EF768CAED458C9 CRC64;

Query Match 100.0%; Score 42; DB 4; Length 1165;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLKDDLLLEA 9
Db 166 VLKDDLLLEA 174

RESULT 2
ID Q9KIV4 PRELIMINARY; PRT; 4150 AA.
AC Q9KIV4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE 8, 8A-DEOXYOLEANOLIDE SYNTHASE 1.
GN OLEA1.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363406; PubMed=10908114;
RA Shah S., Xue Q., Tang L., Carney J.R., Betlach M., McDaniel R.;
RT "Cloning, Characterization, and Heterologous Expression of a
RT polyketide Synthase and P-450 Oxidase Involved in the Biosynthesis of
RT the Antibiotic Oleandomycin.";
RL J. Antibiot. 53:502-508(2000).
DR EMBL; AF220951; AAB82408.1; -.
DR InterPro; IPR000169; -.
DR InterPro; IPR000255; -.
DR InterPro; IPR000794; -.
DR InterPro; IPR001227; -.
DR InterPro; IPR002106; -.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00698; Acyl_transf; 3.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 4150 AA; 435266 MW; 9383296C4C16647D CRC64;

Query Match 90.5%; Score 38; DB 2; Length 4150;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LRDDLLLEA 9
Db 1781 LRDDLLLEA 1788

RESULT 3
ID Q94256 PRELIMINARY; PRT; 174 AA.
AC Q94256;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE COSMID K04A8.
GN K04A8.9;
OS Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Shalton N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Bradshaw H.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64849; AAC48055.1; -.
DR InterPro; IPR000004; -.
DR SMART; SM00118; SAMP; 1.
SQ SEQUENCE 174 AA; 19210 MW; A6303CF6383BED9D9 CRC64;

Query Match 85.7%; Score 36; DB 5; Length 174;
Best Local Similarity 77.8%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLKDDLLLEA 9
Db 129 VLKDDLLLEA 137

RESULT 4
ID Q27025 PRELIMINARY; PRT; 616 AA.
AC Q27025;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE TYPE I RESTRICTION MODIFICATION ENZYME, SUBUNIT M.
GN MTH942.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., DuBois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadaro R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000868; AAB85440.1; -.
DR InterPro; IPR002052; -.
DR InterPro; IPR002296; -.
DR InterPro; IPR003356; -.
DR Pfam; PF02384; N6_Mtase; 1.
DR PRINTS; PR00507; N12NGMTFRASE.

```

DR PROSITE: PS00092; NG\_MTASE: UNKNOWN.1.  
SQ SEQUENCE 616 AA; 71715 MW; 06DAF1076A5D5BAD CRC64;

Query Match 83.3%; Score 35; DB 1; Length 616;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLKDDLLLEA 9  
Db 465 VLEDDLLIEA 473

RESULT 5

Q9WHX7 PRELIMINARY; PRT; 238 AA.

AC Q9WHX7; DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M. influenza A virus (A/Taiwan/96/179).  
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94648;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/96/179;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF138707; AAD33736.2; -  
DR HSSP; P03485; 1AA7.  
DR InterPro; IPR001561; -  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 238 AA; 26266 MW; 401A494B5B673FOA CRC64;

Query Match 81.0%; Score 34; DB 14; Length 238;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 LRDDLLLE 8  
Db 215 LRDDLLLE 221

RESULT 6

Q9WHY9 PRELIMINARY; PRT; 239 AA.

AC Q9WHY9; DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M. influenza A virus (A/Taiwan/98/346).  
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94660;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/98/346;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138719; AAD33768.2; -

DR HSSP; P03485; 1AA7.  
DR InterPro; IPR001561; -  
DR Pfam; PF00598; Flu\_M1; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26462 MW; 2BD3359212691DD8 CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 LRDDLLLE 8  
Db 216 LRDDLLLE 222

RESULT 7

Q9WHY8 PRELIMINARY; PRT; 239 AA.

AC Q9WHY8; DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M. influenza A virus (A/Taiwan/98/224).  
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94659;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/98/224;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138718; AAD33767.2; -  
DR HSSP; P03485; 1AA7.  
DR InterPro; IPR001561; -  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26411 MW; 293533A39097E3E4 CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 LRDDLLLE 8  
Db 216 LRDDLLLE 222

RESULT 8

Q9WHY7 PRELIMINARY; PRT; 239 AA.

AC Q9WHY7; DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M. influenza A virus (A/Taiwan/98/45).  
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94658;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/98/45;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;

RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in  
Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138717; AAD33766.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26381 MW; B24B9FE220299ECD9 CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 9  
Q9WHY6 PRELIMINARY; PRT; 239 AA.  
AC Q9WHY6;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/98/21).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94657;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/98/21;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in  
Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138716; AAD33765.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26298 MW; 6F9A44A4FBFC74BD CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 10  
Q9WHY5 PRELIMINARY; PRT; 239 AA.  
AC Q9WHY5;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/98/20).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.

OX NCBI\_TaxID=94656;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/98/20;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in  
Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138715; AAD33764.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26275 MW; 66345804B4B2C7AC CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 11  
Q9WHY4 PRELIMINARY; PRT; 239 AA.  
AC Q9WHY4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/97/3469).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94655;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/TAIWAN/97/3469;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in  
Taiwan."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138714; AAD33763.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26397 MW; 6E215E154BFC64AC CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 12  
Q9WHY3 PRELIMINARY; PRT; 239 AA.  
AC Q9WHY3;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/97/3351).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94654;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/TAIWAN/97/3351;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138713; AAD33762.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26284 MW; 598A982BB966975A CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 13  
Q9WHY1 ID Q9WHY1 PRELIMINARY; PRT; 239 AA.  
AC Q9WHY1;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/96/2191).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94652;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/TAIWAN/96/2191;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138711; AAD33760.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26439 MW; 8151D148E9968759 CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 14  
Q9WHX9

ID Q9WHX9 PRELIMINARY; PRT; 239 AA.  
AC Q9WHX9;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/96/1600).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94650;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/TAIWAN/96/1600;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138709; AAD33758.2; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 239 AA; 26407 MW; F8CAB96A8644455E CRC64;

Query Match 81.0%; Score 34; DB 14; Length 239;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

RESULT 15  
Q9WHY2 ID Q9WHY2 PRELIMINARY; PRT; 250 AA.  
AC Q9WHY2;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MATRIX PROTEIN M1 (FRAGMENT).  
GN M.  
OS Influenza A virus (A/Taiwan/96/3513).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_TaxID=94653;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/TAIWAN/96/3513;  
RA Shih S.-R., Tsai H.-R., Chang S.-C.;  
RT "Sequence Analysis of M Gene of Influenza A Viruses Isolated in Taiwan";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF138712; AAD33761.1; -.  
DR HSSP; P03485; IAA7.  
DR InterPro; IPR001561; -.  
DR Pfam; PF00598; Flu\_M1; 1.  
DR ProDom; PD001061; -; 1.  
KW Matrix protein.  
FT NON\_TER 1  
SQ SEQUENCE 250 AA; 27671 MW; A47536EC384C3FBC CRC64;

Query Match 81.0%; Score 34; DB 14; Length 250;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
Db 216 LRDDLLE 222

Tue Jul 10 13:30:21 2001

us-09-489-760-5.rspt

Page 6

Db 227 LRDDLE 233

Search completed: July 10, 2001, 08:36:07  
Job time: 260 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:28:11 : Search time 22.85 Seconds  
(without alignments)  
30.003 Million cell updates/sec

Title: US-09-489-760-2

Perfect score: 45

Sequence: 1 VLHDDLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	84.4	1148	2 S51855	hypothetical prote
2	35	77.8	119	2 T00151	hypothetical prote
3	35	77.8	604	2 S36487	Ei protein - human
4	35	77.8	605	2 S36459	Ei protein - human
5	35	77.8	616	2 C69226	type I restriction
6	34	75.6	130	2 A81316	chemotaxis regulat
7	33	73.3	145	2 A41652	probable glyceroph
8	33	73.3	210	2 T44122	probable glyceroph
9	33	73.3	238	2 S76860	hypothetical prote
10	33	73.3	259	2 G82601	conserved hypothet
11	33	73.3	291	2 D83371	probable dehydroge
12	33	73.3	320	2 B75457	polyrenyl synthas
13	33	73.3	322	2 C83075	octaprenyl-dihosp
14	33	73.3	343	2 A45251	DNA-directed DNA p
15	33	73.3	343	2 B95564	DNA polymerase III
16	33	73.3	378	2 T35403	probable polyreny
17	33	73.3	397	2 S90813	hypothetical prote
18	33	73.3	398	2 H72660	probable N-Acylami
19	33	73.3	519	1 S69988	unspecific monooxy
20	33	73.3	519	1 S69989	unspecific monooxy
21	33	73.3	604	2 S36493	Ei protein - human
22	33	73.3	667	2 H85734	probable collagena
23	33	73.3	667	2 F64895	hypothetical prote
24	33	73.3	723	2 F85047	hypothetical prote
25	33	73.3	856	2 C85023	hypothetical prote
26	33	73.3	908	2 T50695	seCA protein [limp
27	33	73.3	973	2 T50449	DNA repair and rec
28	33	73.3	1127	2 T47445	MDM1 protein - yea
29	33	73.3	1770	2 S56221	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S51855  
hypothetical protein YDR128w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YD9302.03

C:Species: Saccharomyces cerevisiae

C>Date: 05-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 26-May-2000

C:Accession: S51855

R:Oliver, K.; Harris, D.

submitted to the EMBL Data Library, February 1995

A:Reference number: S51853

A:Accession: S51855

A:Molecule type: DNA

A:Residues: 1-1148 <OLI>

A:Cross-references: EMBL:Z48179; NID:9665657; PID:9665660; GSPDB:GM000004; MIPS:YDR128

C:Genetics:

A:Gene: MIPS:YDR128w

A:Map position: 4R

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 84.4%; Score 38; DB 2; Length 1148;  
Best Local Similarity 66.7%; Pred. No. 40;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLHDDLEA 9  
:||||:|  
DB 1001 LLHDDIEA 1009

##### RESULT 2

T00151  
hypothetical protein 18 - Staphylococcus aureus phage phi PVL

C:Species: Staphylococcus aureus phage phi PVL

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000

C:Accession: T00151

R:Kaneko, J.; Kimura, T.; Tomita, T.; Kamio, Y.

Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997

A:Title: Pantone-Valentine leukocidin genes in a phage-like particle isolated from ml

A:Reference number: Z14119; MUID:98067870

A:Accession: T00151

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-119 <KAN>

A:Cross-references: EMBL:AB009866; NID:g3341907; PIDN:BAA31891.1; PID:g3341925

C:Superfamily: Staphylococcus aureus phage phi PVL hypothetical protein 18

Query Match 77.8%; Score 35; DB 2; Length 119;  
Best Local Similarity 85.7%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDDLLEA 9

Db 72 HDDILEA 78  
|||||

## RESULT 3

S36487  
EI protein - human papillomavirus type 19

C:Species: human papillomavirus type 19  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999

C:Accession: S36487

R:Delius, H.; Hofmann, B.

submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36487

A:Molecule type: DNA

A:Residues: 1-604 <DEL>

A:Cross-references: EMBL:X74470; NID:g396940; PIDN:CAA52520.1; PID:g396943

C:Superfamily: papillomavirus EI protein

C:Keywords: early protein; nucleus

Query Match 77.8%; Score 35; DB 2; Length 604;

Best Local Similarity 75.0%; Pred. No. 73;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLEA 9

|||||

Db 201 VHDDLES 208

## RESULT 4

S36469  
EI protein - human papillomavirus type 14D

C:Species: human papillomavirus type 14D  
C:Date: 09-Dec-1993 #sequence\_revision 26-Jul-1996 #text\_change 26-Aug-1999

C:Accession: S36469

R:Delius, H.; Hofmann, B.

submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36469

A:Molecule type: DNA

A:Residues: 1-605 <DEL>

A:Cross-references: EMBL:X74467; NID:g396918; PIDN:CAA52502.1; PID:g396919

A:Experimental source: strain 14D

C:Superfamily: papillomavirus EI protein

C:Keywords: early protein; nucleus

Query Match 77.8%; Score 35; DB 2; Length 605;

Best Local Similarity 75.0%; Pred. No. 73;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLEA 9

|||||

Db 202 VHDDLES 209

## RESULT 5

C69226

type I restriction modification enzyme, subunit M - Methanobacterium thermoautotrophicum

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: C69226

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

Qu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct

A:Reference number: A69000; MUID:98037514

A:Accession: C69226

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-616 <MTH>

A:Cross-references: GB:AE000868; GB:AE000666; NID:g2622025; PIDN:AAB85440.1; PID:g2622025

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH942

A:Start codon: GTG

C:Superfamily: type I site-specific deoxyribonuclease chain hsdM

Query Match 77.8%; Score 35; DB 2; Length 616;

Best Local Similarity 77.8%; Pred. No. 74;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLHDDLEA 9

|||||

Db 465 VLEDDLEA 473

## RESULT 6

A81316

chemotaxis regulatory protein Cj1118c [Imported] - Campylobacter jejuni (strain NCTC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000

C:Accession: A81316

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MUID:20150912

A:Accession: A81316

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-130 <PAR>

A:Cross-references: GB:A1139077; GB:A1111168; NID:g6968444; PIDN:CAB73373.1; PID:g6968444

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: cheY; Cj1118c

C:Superfamily: chemotaxis cheY protein; response regulator homology

Query Match 75.6%; Score 34; DB 2; Length 130;

Best Local Similarity 85.7%; Pred. No. 20;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDLLEA 9

|||||

Db 25 HDVLEA 31

## RESULT 7

A41652

probable glycerophosphodiester phosphodiesterase (EC 3.1.4.46) - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 03-Aug-1992 #sequence\_revision 03-Aug-1992 #text\_change 15-Oct-1999

C:Accession: A41652

R:Ryffel, C.; Bucher, R.; Kayser, F.H.; Berger-Baechli, B.

J. Bacteriol. 173, 7416-7422, 1991

A:Title: The Staphylococcus aureus mec determinant comprises an unusual cluster of di

terase.

A:Reference number: A41652; MUID:92041650

A:Accession: A41652

A:Molecule type: DNA

A:Residues: 1-145 <RYF>

A:Cross-references: GB:X52594; NID:g48712; PIDN:CAA36830.1; PID:g48713

C:Keywords: phosphoric diester hydrolase

Query Match 73.3%; Score 33; DB 2; Length 145;

Best Local Similarity 62.5%; Pred. No. 36;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLHDDLE 8





A:Description: Twelve open reading frames revealed on the 23.6 kbp segment flanking the  
 A:Reference number: S45119  
 A:Accession: S45126  
 A:Molecule type: DNA  
 A:Residues: 1-282 <VER>  
 A:Cross-references: EMBL:X77395; NID:g496717; PIDN:CAA54571.1; PID:g496720  
 R:Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.  
 Yeast 10, 1355-1361, 1994  
 A:Title: Twelve open reading frames revealed in the 23.6 kb segment flanking the centrom  
 A:Reference number: S48338; MUID:95208356  
 A:Accession: S48340  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-282 <VE2>  
 A:Cross-references: EMBL:X77395; NID:g496717; PIDN:CAA54571.1; PID:g496720  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1994  
 R:Aert, R.; Verhasselt, P.; Voet, M.; Volckaert, G.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S62910  
 A:Accession: S63328  
 A:Molecule type: DNA  
 A:Residues: 1-282 <AER>  
 A:Cross-references: EMBL:X771617; NID:g1302470; PIDN:CAA96278.1; PID:g1302471; MIPS:YNR0C  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD:FUN34  
 A:Cross-references: SGD:S0005285; MIPS:YNR002c  
 A:Map position: 14R  
 C:Superfamily: Saccharomyces cerevisiae probable membrane protein FUN34  
 C:Keywords: transmembrane protein  
 F:119-135/Domain: transmembrane #status predicted <TM1>  
 F:186-202/Domain: transmembrane #status predicted <TM2>  
 F:209-225/Domain: transmembrane #status predicted <TM3>  
 F:234-250/Domain: transmembrane #status predicted <TM4>

Query Match 81.0%; Score 34; DB 2; Length 282;  
 Best Local Similarity 87.5%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LRDDLLEA 9  
 |||||  
 DB 62 LRDDLFEA 69

RESULT 3  
 A6443  
 glyceralddehyde-3-phosphate dehydrogenase (EC 1.2.1.12) - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: A6443  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
 A:Reference number: A64300; MUID:9637999  
 A:Accession: A6443  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-343 <BU1>  
 A:Cross-references: GB:U67557; GB:L77117; NID:g1591777; PIDN:AAB99147.1; PID:g1591778; T  
 C:Genetics:  
 A:Map position: REV1086336-1085305  
 C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase  
 C:Keywords: oxidoreductase

Query Match 81.0%; Score 34; DB 2; Length 343;  
 Best Local Similarity 77.8%; Pred. No. 38;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDDLLEA 9  
 |||||  
 DB 238 VSRDDLLEA 246

## RESULT 4

T20604  
 hypothetical protein Y51H1A.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T20604; T27102  
 R:Wallis, J.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19298  
 A:Accession: T20604  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-389 <WIL>  
 A:Cross-references: EMBL:Z81495; PIDN:CA804064.1; GSPDB:GN000020; CESP:Y51H1A.1  
 A:Experimental source: clone F08G2  
 R:Smeye, R.  
 submitted to the EMBL Data Library, October 1998  
 A:Reference number: Z20309  
 A:Accession: T27102  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-389 <W12>  
 A:Cross-references: EMBL:AL032644; PIDN:CAA21673.1; GSPDB:GN000020; CESP:Y51H1A.1  
 A:Experimental source: clone Y51H1A  
 C:Genetics:  
 A:Gene: CESP:Y51H1A.1  
 A:Map position: 2  
 A:Introns: 12/3; 38/3; 167/2; 276/1; 342/3

Query Match 81.0%; Score 34; DB 2; Length 389;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLEA 9  
 |||||  
 DB 21 LRDELLEA 28

## RESULT 5

A82878  
 replicative DNA helicase U0550 [imported] - Ureaplasma urealyticum  
 C:Species: Ureaplasma urealyticum  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: A82878  
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
 submitted to GenBank, February 2000  
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a  
 A:Reference number: A82870  
 A:Accession: A82878  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-486 <GLA>  
 A:Cross-references: GB:AF002153; GB:AF222894; NID:g6899544; PIDN:AAF30963.1; GSPDB:GN  
 A:Experimental source: serovar 3; biovar 1  
 C:Genetics:  
 A:Gene: dnab; U0550  
 A:Genetic code: SGC3

Query Match 81.0%; Score 34; DB 2; Length 486;  
 Best Local Similarity 75.0%; Pred. No. 56;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLEA 9  
 :|||||  
 DB 445 MRDDLLEA 452

A:Title: Evolutionary conservation of the EPS8 gene and its mapping to human chromosome  
 A:Reference number: I38728; MUID:94366758  
 A:Accession: I38728  
 A:Status: preliminary; translated from GB/EMBL/DBSJ  
 A:Molecule type: mRNA  
 A:Residues: 1-822 <RES>  
 A:Cross-references: EMBL:U12535; NID:g530822; PID:g530823  
 C:Genetics:  
 A:Gene: Eps8  
 C:Superfamily: SH3 homology

Query Match 78.6%; Score 33; DB 2; Length 822;  
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLKDDILE 8  
 |||||  
 Db 552 VLKDDILE 559

## RESULT 11

E70760  
 probable lles protein - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: E70760  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: E70760  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1041 <COL>  
 A:Cross-references: GB:Z74020; GB:AL123456; NID:g3261584; PIDN:CAA98326.1; PID:g1403506  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: lles  
 C:Superfamily: isoleucine--trna ligase

Query Match 78.6%; Score 33; DB 2; Length 1041;  
 Best Local Similarity 77.8%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLKDDILE 9  
 |||||  
 Db 710 VLKDDILE 718

## RESULT 12

D70032  
 two-component response regulator [YvcQ] homolog yvcP - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: D70032  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin  
 A:Authors: Laubet, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
 Rieger, A.M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon  
 A:Authors: Schleith, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, M

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: D70032  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-237 <KUN>  
 A:Cross-references: GB:Z99121; GB:AL009136; NID:g2635827; PIDN:CAB15477.1; PID:g26359  
 C:Genetics:  
 A:Gene: yvcP  
 C:Superfamily: ompR protein; response regulator homology  
 C:Keywords: phosphoprotein  
 F:4-112/Domain: response regulator homology <RRH>  
 F:52/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 76.2%; Score 32; DB 2; Length 237;  
 Best Local Similarity 77.8%; Pred. No. 63;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLKDDILE 9  
 |||||  
 Db 176 VSRDELLEA 184

## RESULT 13

T41578  
 hypothetical protein SPC737.05 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Feb-2000  
 C:Accession: T41578  
 R:Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, September 1998  
 A:Reference number: 22002  
 A:Accession: T41578  
 A:Status: translated from GB/EMBL/DBSJ  
 A:Molecule type: DNA  
 A:Residues: 1-264 <MUR>  
 A:Cross-references: EMBL:AL031546; NID:el375214; PIDN:CAB44773.1; GSPDB:GN00068; SPDB  
 A:Experimental source: strain 972h(-)  
 C:Genetics:  
 A:Gene: SPC737.05  
 A:Map position: 3

Query Match 76.2%; Score 32; DB 2; Length 264;  
 Best Local Similarity 55.6%; Pred. No. 71;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLKDDILE 9  
 |||||  
 Db 26 VLKDDILE 34

## RESULT 14

A64988  
 hypothetical 37.8 kD protein in rply-proL intergenic region - Escherichia coli (strai  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
 C:Accession: A64988  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: A64988  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-335 <BLAT>  
 A:Cross-references: GB:AB000308; GB:U000096; NID:gl788508; PIDN:AAC75247.1; PID:gl788;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:

```

QY      1 VLRDDLLLE 8
      |:||||:|
Db      144 VLRLDLVE 151

RESULT      8
Tl1672
quinone oxidoreductase homolog - cowpea
C:Species: Vigna unguiculata (cowpea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: Tl1672
R:Krause, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z17311
A:Accession: Tl1672
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-324 <KRA>
A:Cross-references: EMBL:Y08624
A:Experimental source: cultivar red caloona; root hair
C:Genetics:
A:Gene: Red2
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
F:26-314/Domain: long-chain alcohol dehydrogenase homology <LAD>

```

```

Query Match          78.6%; Score 33; DB 2; Length 324;
Best Local Similarity 77.8%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 VLRDDELEA 9
        | |:||||
Db      270 VTRDELEA 278

RESULT          9
H85073

```

C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: H85073  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The C  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana  
A;Reference number: AB5001; MUID:20083488  
A;Accession: H85073

```

A:MoreInfo type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:NC_001268; NID:g7267351; PIDN:CAB81124.1; GSPDB:GNO0140
C:Genetics:
A:Gene: AY4907540
A:Map position: 4

Query Match          78.6%; Score 33; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps. 0;

QY      1 VLRDDL 7
          | | | | |
Db      145 VLRDDL 151

RESULT 10
I38728
epidermal growth factor receptor kinase substrate - human

```

C; Date: 29-May-1998 #sequence\_revision 29-May-1998 #tex\_change 09-Jun-2000  
C; Accession: 138728  
R; Wong, W.T.; Carlomagno, F.; Druck, T.; Barletta, C.; Croce, C.M.; Huebner, K.  
Oncogene 9, 3057-3061, 1994

DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE C-TERMINAL SEQUENCE OF ORF C PRODUCT SHOWS HOMOLOGU TO SEVERAL  
 DE BACTERIAL ORFS.  
 OS Leishmania donovani.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MHOM/BL/67/ITMAP263;  
 RX MEDLINE=95047425; PubMed=7958944;  
 RA Myler P.J., Venkataranan G.M., Lodes M.J., Stuart K.D.;  
 RT "A frequently amplified region in Leishmania contains a gene conserved  
 in prokaryotes and eukaryotes."  
 RL Gene 148:187-193(1994).  
 DR EMBL; U02459; AAA64631.1; -.  
 DR InterPro; IPR000613; -.  
 DR Pfam; PF00849; Pseudou\_synth\_2; 1.  
 DR ProDom; PD001819; -; 1.  
 DR PROSITE; PS01129; PSI\_RLU; 1.  
 SQ SEQUENCE 741 AA; 80060 MW; AE28C60ABDFC6C72 CRC64;  
 Query Match 75.6%; Score 34; DB 5; Length 741;  
 Best Local Similarity 62.5%; Pred. No. 3e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 LHDDLEA 9  
 DB 628 LHDDIIDA 635  
 RESULT 9  
 Q9NDK7 PRELIMINARY; PRT; 741 AA.  
 AC Q9NDK7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE 511.8.  
 GN Leishmania major.  
 OS Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Myler P.J., Sisk E., Hixson G., Kiser P., Rickel E., Hassebrock M.,  
 RA Cavthra J., Marsolini F., Sunkin S., Stuart K.D., Cunningham M.,  
 RA Beverley S.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005801; AAF77202.1; -.  
 DR InterPro; IPR000345; -.  
 DR InterPro; IPR000613; -.  
 DR InterPro; IPR002990; -.  
 DR Pfam; PF00849; Pseudou\_synth\_2; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN1.  
 DR PROSITE; PS01129; PSI\_RLU; 1.  
 SQ SEQUENCE 741 AA; 79890 MW; 81AE049B109A3F2 CRC64;

Query Match 75.6%; Score 34; DB 5; Length 741;  
 Best Local Similarity 62.5%; Pred. No. 3e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLEA 9  
 DB 628 LHDDIIDA 635

RESULT 10.

Q9VA13 PRELIMINARY; PRT; 774 AA.  
 ID Q9VA13;  
 AC Q9VA13;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CG11318 PROTEIN.  
 DE CG11318.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2183-2195(2000).  
 DR EMBL; AE003776; AAF57115.1; -.  
 DR FlyBase; FBgn0039818; CG11318.  
 DR InterPro; IPR000203; -.  
 DR InterPro; IPR000832; -.  
 DR Pfam; PF00002; 7tm.2; 1.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00303; GPS; 1.  
 SQ SEQUENCE 774 AA; 87359 MW; 0C3FF3A99A12CF79 CRC64;

Query Match 75.6%; Score 34; DB 5; Length 774;  
 Best Local Similarity 85.7%; Pred. No. 3.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLE 8  
 DB 121 LHDDVIE 127

RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,  
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,  
 RA "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL: AE000868; AAB85440.1; -  
 DR InterPro: IPR002052; -  
 DR InterPro: IPR003356; -  
 DR Pfam: PF02384; N6\_Mtase; 1.  
 DR PRINTS: PR00507; N12N6WTFRASE.  
 DR PROSITE: PS00092; NO\_MTASE; UNKNOWN\_1.  
 SQ SEQUENCE 616 AA; 71715 MW; 06D4F1076A5D58AD CRC64;

Query Match 77.8%; Score 35; DB 1; Length 616;  
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLHDDLLEA 9  
 |||||  
 DB 465 VLEDDLLEA 473

RESULT 6  
 ID Q9V966 PRELIMINARY; PRT; 332 AA.  
 AC Q9V966;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE CG9025 PROTEIN.  
 GN CG9025.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.F., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrara C., Ferrera S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacliet J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan K.C., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003791; AAF57431.1; -  
 DR HSSP: P55273; 16D8  
 DR FlyBase: FBgn0034542; CG9025.  
 DR InterPro: IPR002110; -  
 DR Pfam: PF00023; ank; 5.  
 DR PROSITE: PS50088; ANK\_REPEAT; 3.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR SMART: SM00248; ANK; 1.  
 SQ SEQUENCE 332 AA; 36359 MW; DD6F2CD87C66186E CRC64;

Query Match 75.6%; Score 34; DB 5; Length 332;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLLEA 9  
 |||||  
 DB 285 LHEDVLEA 292

RESULT 7  
 ID Q9NKM2 PRELIMINARY; PRT; 640 AA.  
 AC Q9NKM2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE 511.8 (FRAGMENT).  
 GN 511.8.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Myler P.J.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC009603; AAF70540.1; -  
 DR InterPro: IPR000345; -  
 DR InterPro: IPR000613; -  
 DR InterPro: IPR002990; -  
 DR Pfam: PF00849; Pseudou-synth\_2; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE: PS01129; PSI\_RLU; 1.  
 FT NON\_TER 640  
 SQ SEQUENCE 640 AA; 68503 MW; 8A19583709E1BB7D CRC64;

Query Match 75.6%; Score 34; DB 5; Length 640;  
 Best Local Similarity 62.5%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLLEA 9  
 |||||  
 DB 628 LHDDIIDA 635

RESULT 8  
 ID Q25257 PRELIMINARY; PRT; 741 AA.  
 AC Q25257;

Query Match 75.6%; Score 34; DB 5; Length 1383;

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
 RA Craxton M., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
 RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
 RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans\*;  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Geisel C., Bradshaw H.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U64849; AAC48055.1; -;  
 DR InterPro: IPR000004; -;  
 DR SMART: SM00118; SAPB; 1;  
 SQ SEQUENCE 174 AA; 19210 MW; A6303CF6383BEBD9 CRC64;

Query Match 73.3%; Score 33; DB 5; Length 174;  
 Best Local Similarity 77.8%; Pred. NO. le+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLHDDLLEA 9  
 || |||||  
 Db 129 VLKDDLLDA 137

RESULT 14  
 Q9RPT9 PRELIMINARY; PRT; 175 AA.  
 ID Q9RPT9;  
 AC Q9RPT9;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMELrel. 14, Last annotation update)  
 DE HYPOTHETICAL 20.3 KDA PROTEIN.  
 OS Streptomyces albus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC21838;  
 RA Kwon H.-J., Lee S.-Y., Hong S.-K., Park U.-M., Suh J.-W.;  
 RT "Heterologous Expression of Streptomyces albus Genes Linked to an  
 Integrating Element and Activation of Antibiotic Production.";  
 RL J. Microbiol. Biotechnol. 9:488-497(1999).  
 DR EMBL: AF145724; AAD46513.1; -;  
 DR InterPro: IPR000182; -;  
 DR Pfam: PF00583; Acetyltransf; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 175 AA; 20300 MW; 22A46EA40732DF0A CRC64;

Query Match 73.3%; Score 33; DB 2; Length 175;  
 Best Local Similarity 100.0%; Pred. NO. le+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDLLE 8  
 |||||  
 Db 33 HDLLE 18

RESULT 15  
 Q9S3K5 PRELIMINARY; PRT; 210 AA.  
 ID Q9S3K5;  
 AC Q9S3K5;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DE GLYCEROPHOSPHORYLDIESTER PHOSPHODIESTERASE (FRAGMENT).  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPOSON-TN55X;  
 RX MEDLINE=99278010; PubMed=10348769;  
 RA Ito T., Katayama Y., Hiramatsu K.;  
 RT "cloning and nucleotide sequence determination of the entire mec DNA  
 of pre-methicillin-resistant Staphylococcus aureus N315.";  
 RL Antimicrob. Agents Chemother. 43:1449-1458(1999).  
 DR EMBL: D86934; BAA82224.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 210 AA; 23757 MW; 3F0B88A13397D668 CRC64;

Query Match 73.3%; Score 33; DB 2; Length 210;  
 Best Local Similarity 62.5%; Pred. NO. 1.3e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLHDDLLE 8  
 :||| |  
 Db 48 ITHDDVLE 55

Search completed: July 10, 2001, 08:36:07  
 Job time: 260 sec



A:Gene: yejK

Query Match 76.2%; Score 32; DB 2; Length 335;  
Best Local Similarity 87.5%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDLDLLE 8  
      |||||  
DB 24 VLRDLSLLE 31

RESULT 15

H85857  
hypothetical protein yejK [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: H85857  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: H85857  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-335 <STO>  
A:Cross-references: GB:AE005174; NID:2516518; PIDN:AAG57324.1; GSPDB:GN00145; UWGP:234  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: yejK

Query Match 76.2%; Score 32; DB 2; Length 335;  
Best Local Similarity 87.5%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDLDLLE 8  
      |||||  
DB 24 VLRDLSLLE 31

Search completed: July 10, 2001, 08:31:41  
Job time: 210 sec

7

2.8



Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: B75457

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-320 <WHI>

A:Cross-references: GB:A8001946; GB:A8000513; NID:g6458655; PIDN:AAF10509.1; PID:g645865

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0932

A:Map position: 1

C:Superfamily: prenyl transferase A

Query Match 73.3%; Score 33; DB 2; Length 320;

Best Local Similarity 62.5%; Pred. No. 86;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLHDDLL 8

:|||||:

Db 79 LLHDDLLD 86

RESULT 13

C83075

Occiprenyl-diphosphate synthase PA4569 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83075

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: C83075

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-322 <STO>

A:Cross-references: GB:A8004870; GB:A8004091; NID:g9950807; PIDN:AA07957.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: ispB; PA4569

C:Superfamily: prenyl transferase A

Query Match

Best Local Similarity 73.3%; Score 33; DB 2; Length 322;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLHDDLL 9

:|||||:

Db 81 LLHDDVVDA 89

RESULT 14

A45251

DNA-directed DNA polymerase (EC 2.7.7.7) III delta chain - Escherichia coli

C:Species: Escherichia coli

C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999

C:Accession: A45251; A46738; F64798

R:Carter, J.R.; Franden, M.A.; Aebersold, R.; McHenry, C.S.

J. Bacteriol. 174, 7013-7025, 1992

A:Title: Molecular cloning, sequencing, and overexpression of the structural gene encodi

A:Reference number: A45251; MUID:93015766

A:Accession: A45251

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-343 <CAR>

A:Cross-references: GB:M94267; NID:g145784; PIDN:AAB59047.1; PID:g145785

A:Note: sequence extracted from NCBI backbone (NCBIP:117007)

R:Dong, Z.; Onrust, R.; Skangalis, M.; O'Donnell, M.

J. Biol. Chem. 268, 11758-11765, 1993

A:Title: DNA polymerase III accessory proteins. I. holoA and holoB encoding delta and

A:Reference number: A46738; MUID:93280136

A:Accession: A46738

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-343 <DON>

A:Cross-references: GB:L04576; NID:g145728; PIDN:AAA23675.1; PID:g145729

A:Experimental source: strain K-12

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.,

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64798

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-343 <BLAT>

A:Cross-references: GB:A8000168; GB:U00096; NID:g1786849; PIDN:AAC73741.1; PID:g17868

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: holoA

C:Complex: DNA polymerase III is a multichain complex; alpha, epsilon, theta, tau, ga

C:Function: <GEN>

A:Description: DNA synthesis; synthesizes both, the lagging and the leading strands 1

A:Pathway: DNA biosynthesis

A:Note: core enzyme (catalytic core) contains chains alpha, epsilon and theta; alpha

required for assembly; tau allows dimerization of the core complex and processivity

te; the holoenzyme is completed by addition of beta chain which clamps the enzyme to

C:Function: <DEL>

A:Description: delta chain seems to interact with the gamma chain

A:Note: gamma/delta complex is composed of chains gamma, delta, delta', psi and chi

C:Keywords: DNA biosynthesis; DNA replication; nucleotidyltransferase

Query Match

Best Local Similarity 73.3%; Score 33; DB 2; Length 343;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLHDDLL 7

:|||||:

Db 103 LLHDDLL 109

RESULT 15

B85564

DNA polymerase III, delta subunit [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: B85564

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: B85564

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-343 <STO>

A:Cross-references: GB:A8005174; NID:g12513544; PIDN:AAG54974.1; GSPDB:GN00145; UWGP

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: holoA

Query Match

Best Local Similarity 73.3%; Score 33; DB 2; Length 343;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLHDDLL 7

:|||||:

Db 103 LLHDDLL 109

us-09-489-760-2.rpr

Tue Jul 10 13:30:14 2001

Search completed: July 10, 2001, 08:31:41  
Job time: 210 sec

7



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:31:17 ; Search time 12.64 Seconds  
(without alignments)  
24.391 Million cell updates/sec

Title: US-09-489-760-2

Perfect score: 45

Sequence: 1 VLHDDLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	35	77.8	603	1	VEL_HPV21	P50759	human papill
2	35	77.8	604	1	VEL_HPV19	Q02048	human papill
3	35	77.8	605	1	VEL_HPV14	P36721	human papill
4	34	75.6	130	1	CHEY_CAMJE	P71129	campylobact
5	33	73.3	343	1	HOLA_ECOLI	P28630	escherichia
6	33	73.3	397	1	UL50_HCMVA	P16791	human cytom
7	33	73.3	443	1	MDM1_YEAST	Q01846	saccharomyc
8	33	73.3	519	1	CP5W_CANAP	Q33083	candida api
9	33	73.3	519	1	CP5W_CANAP	Q12573	candida api
10	33	73.3	604	1	VEL_HPV25	Q02049	human papill
11	33	73.3	653	1	YDCP_ECOLI	P76104	escherichia
12	33	73.3	1770	1	R115_YEAST	P43565	saccharomyc
13	32	71.1	107	1	RLAL_CHLRE	P29763	chlamydomon
14	32	71.1	166	1	YVAE_VACCC	P20514	vaccinia vi
15	32	71.1	245	1	B2LF_EBV	P03206	epstein-bar
16	32	71.1	284	1	THT3_MYCTU	Q59570	mycobacteri
17	32	71.1	296	1	Y842_CAEEL	Q09371	caenorhabdi
18	32	71.1	410	1	YCAQ_ECOLI	P75843	escherichia
19	32	71.1	736	1	KPCE_RABIT	P10830	oryctolagus
20	32	71.1	737	1	KPCE_HUMAN	Q02156	homo sapien
21	32	71.1	737	1	KPCE_MOUSE	P16054	mus musculus
22	32	71.1	737	1	KPCE_RAT	P09216	rattus norv
23	32	71.1	743	1	KPCE2_APLCA	Q16975	aplysia cal
24	32	71.1	767	1	DLG4_HUMAN	P78352	homo sapien
25	31.5	70.0	549	1	YQCE_ECOLI	P32703	escherichia
26	31	68.9	103	1	RLAL_POLPE	P27464	polyorchis
27	31	68.9	149	1	LRP6_BACSU	P96653	bacillus su
28	31	68.9	231	1	YHHW_ECOLI	P46852	escherichia
29	31	68.9	247	1	UGPO_ECOLI	P10908	escherichia
30	31	68.9	290	1	ARYL_HUMAN	P18440	homo sapien
31	31	68.9	428	1	DGTP_MYCSM	Q52199	mycobacteri
32	31	68.9	469	1	K64_HUMAN	Q92504	homo sapien
33	31	68.9	499	1	AMPA_CHLPN	Q92818	chlamydia p

34 31 68.9 584 1 HEMA\_INBLE P03460 influenza b  
35 31 68.9 586 1 VDMR\_BPT3 P10310 bacterioph  
36 31 68.9 604 1 VEL\_HPV12 Q05134 human papil  
37 31 68.9 604 1 VEL\_HPV36 P50808 human papil  
38 31 68.9 605 1 VEL\_HPV20 P50758 human papil  
39 31 68.9 605 1 VEL\_HPV47 P22419 human papil  
40 31 68.9 606 1 VEL\_HPV05 P06920 human papil  
41 31 68.9 606 1 VEL\_HPV5B P26542 human papil  
42 31 68.9 606 1 VEL\_HPV7 O40619 human papil  
43 31 68.9 688 1 YQAQ\_SCHPO Q10109 schizosacch  
44 31 68.9 727 1 KDGA\_RAT P51556 rattus norv  
45 31 68.9 730 1 KDGA\_MOUSE O88673 mus musculu

## ALIGNMENTS

### RESULT 1

VEL\_HPV21  
ID VEL\_HPV21 STANDARD; PRT; 603 AA.  
AC P50759;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE REPLICATION PROTEIN E1.  
GN E1.  
OS Human papillomavirus type 21.  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
OX NCBI\_TaxID=31548;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Delius H.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF  
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2  
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH  
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: U31779; AAA79396.1; -  
CC InterPro: IPR001177; -  
CC Pfam: PF00519; E1; 1.  
CC Pfam: PF00524; E1\_N; 1.  
CC Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;  
CC Nuclear protein.  
CC NP\_BIND 431 438  
CC FT NP\_BIND 431 438  
CC SEQUENCE 603 AA; 69983 MW; 93006494BEF10339 CRC64;

Query Match 77.8%; Score 35; DB 1; Length 603;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LHHDDLEA 9  
:|||||:  
Db 200 VHHDDLES 207

### RESULT 2

VEL\_HPV19  
ID VEL\_HPV19 STANDARD; PRT; 604 AA.  
AC Q02048;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE GN REPLICATION PROTEIN E1.  
 OS Human papillomavirus type 19.  
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
 OX NCBI\_TaxID=10608;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94265501; PubMed=8205838;  
 RA Delius H., Hofmann B.;  
 RA "Primer-directed sequencing of human papillomavirus types.";  
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).  
 RN [2]  
 RP SEQUENCE OF 331-382 FROM N.A.  
 RX MEDLINE=92407963; PubMed=1326639;  
 RA Chan S.Y., Bernard H.U., Ong C.K., Chan S.P., Birgit H., Delius H.;  
 RA "Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and  
 RT variants: a showcase for the molecular evolution of DNA viruses.";  
 RL J. Virol. 66:5714-5725(1992).  
 CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF  
 CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2  
 CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH  
 CC CONTAINS BINDING SITES FOR BOTH PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X74470; CAA52520.1; -;  
 DR EMBL; M96320; AAA47009.1; -;  
 DR PIR; S36487; S36487;  
 DR InterPro: IPR001177; -;  
 DR Pfam; PF00519; E1; 1.  
 DR Pfam; PF00524; E1\_N; 1.  
 DR Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;  
 KW Nuclear protein.  
 FT NP\_BIND 432 439 ATP (POTENTIAL).  
 SQ SEQUENCE 604 AA; 69153 MW; 6CC7376D631D3FA1 CRC64;

Query Match 77.8%; Score 35; DB 1; Length 604;  
 Best Local Similarity 75.0%; Pred. No. 23;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 LHDDLEA 9  
 DB 201 VHDDLES 208  
 :|||||:  
 201 VHDDLES 208

RESULT 3  
 VEL\_HPV14  
 ID VEL\_HPV14 STANDARD; PRT; 605 AA.  
 AC P36721.  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE REPLICATION PROTEIN E1.  
 DE GN E1.  
 OS Human papillomavirus type 14.  
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
 OX NCBI\_TaxID=10605;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94265501; PubMed=8205838;  
 RA Delius H., Hofmann B.;  
 RA "Primer-directed sequencing of human papillomavirus types.";  
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).  
 CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF  
 CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2

CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH  
 CC CONTAINS BINDING SITES FOR BOTH PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X74467; CAA52502.1; -;  
 DR PIR; S36469; S36469.  
 DR InterPro: IPR001177; -;  
 DR Pfam; PF00519; E1; 1.  
 DR Pfam; PF00524; E1\_N; 1.  
 DR Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;  
 KW Nuclear protein.  
 FT NP\_BIND 433 440 ATP (POTENTIAL).  
 SQ SEQUENCE 605 AA; 69302 MW; 6702915505AA6491 CRC64;

Query Match 77.8%; Score 35; DB 1; Length 605;  
 Best Local Similarity 75.0%; Pred. No. 23;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 LHDDLEA 9  
 DB 202 VHDDLES 209  
 :|||||:  
 202 VHDDLES 209

RESULT 4  
 CHEY\_CAMJE  
 ID CHEY\_CAMJE STANDARD; PRT; 130 AA.  
 AC P71129.  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE CHEMOTAXIS PROTEIN CHEY HOMOLOG.  
 GN CHEY OR CJ1118C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11828;  
 RA Marchant J.E., Henderson J., Wren B.W., Ketley J.M.;  
 RA Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=81116;  
 RA MEDLINE=98386495; PubMed=9720026;  
 RA Fry B.N., Korolik V., ten Brinke J.A., Pennings M.T.T., Zalm R.,  
 RA Teunis B.J.J., Coloe P.J., van der Zeijst B.A.M.;  
 RA "The lipopolysaccharide biosynthesis locus of Campylobacter jejuni  
 RT 81116.";  
 RL Microbiology 144:2049-2061(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RA MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
 RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RL reveals hypervariable sequences.";  
 RN Nature 403:665-668(2000).  
 RN [4]  
 RP SEQUENCE OF 12-130 FROM N.A.



```

RC STRAIN-ATCC 55026 / 81-176;
RX MEDLINE-97231343; PubMed-6738;
RA Yoo R., Burr D.H., Guerry P.;
RT "Chey-mediated modulation of Campylobacter jejuni virulence.";
RL MOL. Microbiol. 23:1021-1032(1997).
CC -1- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM
CC THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEY SEEMS TO REGULATE
CC THE CLOCKWISE (CW) ROTATION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RESPONSE REGULATORY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75208; AAB17571.1; -
DR EMBL: Y11648; CAA72347.1; -
DR EMBL: AL139077; CAB73373.1; -
DR EMBL: U62038; AAC44858.1; -
DR HSSP: P06143; ICEY
DR InterPro: IPR001789; -
DR Pfam: PF00072; response_reg; 1.
KW Chemotaxis; Sensory transduction; Phosphorylation; Flagellar rotation.
FT DOMAIN 1 119
FT MOD_RES 53 53 PHOSPHORYLATION (BY SIMILARITY).
FT SEQUENCE 130 AA; 14437 MW; 41634D8ADE6C5081 CRC64;

Query Match 75.68; Score 34; DB 1; Length 130;
Best Local Similarity 85.78; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDDLLEA 9
| | | | |
DB 25 HDDVLEA 31

RESULT 5
HOLA_ECOLI
ID HOLA_ECOLI STANDARD; PRT; 343 AA.
AC P28630;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE "DNA POLYMERASE III, DELTA SUBUNIT (EC 2.7.7.7)".
GN HOLA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-93280136; PubMed-8505303;
RA Dong Z., Onrust R., Skangalis M., O'Donnell M.;
RT "DNA polymerase III accessory proteins. I. hola and holaB encoding
RT delta and delta'.";
RL J. Biol. Chem. 268:11758-11765(1993).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-WAF102;
RX MEDLINE-93015766; PubMed-1400251;
RA Carter J.R., Franden M.A., Aebersold R.H., McHenry C.S.;
RT "Molecular cloning, sequencing, and overexpression of the structural
RT gene encoding the delta subunit of Escherichia coli DNA polymerase
RT III holoenzyme.";
RL J. Bacteriol. 174:7013-7025(1992).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [6]
RP PRELIMINARY SEQUENCE OF 1-49 FROM N.A.
RC STRAIN-K12;
RX MEDLINE-88058785; PubMed-3316191;
RA Takase I., Ishino F., Wachi M., Kamata H., Doi M., Asoh S.,
RA Matsuzawa H., Ohta T., Matsubashi M.;
RT "Genes encoding two lipoproteins in the leuS-dacA region of the
RT Escherichia coli chromosome.";
RL J. Bacteriol. 169:5692-5699(1987).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE-93280137; PubMed-8505304;
RA Onrust R., O'Donnell M.;
RT "DNA polymerase III accessory proteins. II. Characterization of delta
RT and delta'.";
RL J. Biol. Chem. 268:11766-11772(1993).
RN [8]
RP REVIEW.
RX MEDLINE-92246902; PubMed-1575709;
RA O'Donnell M.;
RT "Accessory protein function in the DNA polymerase III holoenzyme from
RT E. coli.";
RL Bioessays 14:105-111(1992).
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE DELTA SUBUNIT SEEMS TO INTERACT WITH THE GAMMA SUBUNIT TO
CC TRANSFER THE BETA SUBUNIT ON THE DNA.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLII' COMPLEX. POLIII' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN. THE FINAL COMPOSITION OF THE
CC COMPLEX IS: (ALPHA,EPSILON,THETA)[2]-TAU[2]-(GAMMA,DELTA,DELTA',
CC PSI,CHI)[2]-BETA[4].
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; L04576; AAA23675.1; -.
DR EMBL; M94267; AAB59047.1; -.
DR EMBL; AE000168; AAC73741.1; -.
DR EMBL; U82598; AAB40841.1; -.
DR EMBL; D90704; BAA35287.1; -.
DR EMBL; M18277; -. NOT_ANNOTATED_CDS.
DR PIR; A45251; A45251.
DR PIR; A46738; A46738.
DR ECO2DBASE; F032.4; 6TH EDITION.
DR Ecogene; EG11412; hola.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 343 AA; 38703 MW; 87239A678FE16BE2 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 343;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLHDDL 7
DB 103 LLHDDL 109

RESULT 6
UL50_HCMVA STANDARD; PRT; 397 AA.
ID UL50_HCMVA
AC P16791;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE PROTEIN UL50.
GN UL50.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Reddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL Curr Top Microbiol Immunol. 154:125-169(1990).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL34,
CC EHVS-1 26, EBV BFRL1, HVS-1 67, HCMV UL50, AND VZV 24.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; CAA35409.1; -.
DR PIR; S09813;
SQ SEQUENCE 397 AA; 42901 MW; 08DC5FA358E8A473 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 397;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLHDDLLEA 9
DB 6 VLHDDLQVA 14

RESULT 7
MDML_YEAST STANDARD; PRT; 443 AA.
ID MDML_YEAST

```

```

AC Q01846;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE STRUCTURAL PROTEIN MDML.
GN MDML OR YML104C OR YMB339.15C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92332595; PubMed=1378448;
RA McConnell S.J., Yaffe M.P.;
RT "Nuclear and mitochondrial inheritance in yeast depends on novel
RT cytoplasmic structures defined by the MDML protein."
RL J. Cell Biol. 118:385-395(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 190-443 FROM N.A.
RA Stirling C.J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL FOR MITOTIC GROWTH. MEDIANTE ORGANELLE
CC INHERITANCE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: LOW, TO VIMENTIN AND OTHER INTERMEDIATE FILAMENTS.
CC -!- SIMILARITY: TO S.POMBE SPC16A11.04.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66371; CAA47014.1; -.
DR EMBL; Z49210; CAA89114.1; -.
DR EMBL; X65783; CAA46664.1; -.
DR PIR; A42636; A42636.
DR SGD; S0004572; MDML.
DR InterPro; IPR001683; -.
DR Pfam; PF00787; PX; 1.
FT CONFLICT 88 88 Q -> R (IN REF. 1).
FT CONFLICT 215 215 T -> I (IN REF. 1).
FT CONFLICT 379 379 R -> K (IN REF. 1).
FT CONFLICT 397 397 T -> S (IN REF. 1).
FT CONFLICT 412 412 G -> R (IN REF. 1).
SQ SEQUENCE 443 AA; 51470 MW; 9EAA945C8DA70230 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 443;
Best Local Similarity 71.4%; Pred. No. 39;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHDDLLE 8
DB 227 MHDDILE 233

RESULT 8
CP5V_CANAP STANDARD; PRT; 519 AA.
ID CP5V_CANAP
AC P43083;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 52E1 (EC 1.14.14.1) (CYPLIIE1).
GN CYP52E1.

```

OS Candida apicola (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=29830;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IMET 43747;  
 RX MEDLINE=96367597; PubMed=8771711;  
 RA Lottermoser K., Schunck W.H., Asperger O.;  
 RT "Cytochromes P450 of the sophorose lipid-producing yeast Candida  
 RT apicola: heterogeneity and polymerase chain reaction-mediated cloning  
 RT of two genes";  
 RL Yeast 12:565-575(1996).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X76225; CAA53811.1; -  
 DR InterPro; IPR001128; -  
 DR InterPro; IPR002402; -  
 DR InterPro; IPR002974; -  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00464; EP450II.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Electron transport; Oxidoreductase; Monooxygenase; Heme;  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 44 64 POTENTIAL.  
 FT BINDING 479 479 HEME (BY SIMILARITY).  
 SQ SEQUENCE 519 AA; 58656 MW; 4185235A07EA4370 CRC64;  
 Query Match 73.3%; Score 33; DB 1; Length 519;  
 Best Local Similarity 75.0%; Pred. No. 47;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLHDDLLLE 8  
 DB 63 MLHDDNLE 70  
 RESULT 9  
 CP5W\_CANAP STANDARD; PRT; 519 AA.  
 AC Q12573;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CYTOCHROME P450 52E2 (EC 1.14.14.1) (CYPLIIE2).  
 GN CYP52E2.  
 OS Candida apicola (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=29830;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IMET 43747;  
 RX MEDLINE=96367597; PubMed=8771711;  
 RA Lottermoser K., Schunck W.H., Asperger O.;  
 RT "Cytochromes P450 of the sophorose lipid-producing yeast Candida  
 RT apicola: heterogeneity and polymerase chain reaction-mediated cloning  
 RT of two genes";  
 RL Yeast 12:565-575(1996).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X87640; CAA60980.1; -  
 DR InterPro; IPR001128; -  
 DR InterPro; IPR002402; -  
 DR InterPro; IPR002974; -  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00464; EP450II.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Electron transport; Oxidoreductase; Monooxygenase; Heme;  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 44 64 POTENTIAL.  
 FT BINDING 461 461 HEME (BY SIMILARITY).  
 SQ SEQUENCE 519 AA; 58631 MW; F0B164E22D169C86 CRC64;  
 Query Match 73.3%; Score 33; DB 1; Length 519;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HDDLLE 8  
 DB 65 HDDLLE 70  
 RESULT 10  
 VEI\_HPV25 STANDARD; PRT; 604 AA.  
 ID VEI\_HPV25  
 AC Q02049;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE REPLICATION PROTEIN E1.  
 GN E1.  
 OS Human papillomavirus type 25.  
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
 OX NCBI\_TaxID=10609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94265501; PubMed=8205838;  
 RA Delius H., Hofmann B.;  
 RT "Primer-directed sequencing of human papillomavirus types";  
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).  
 RN [2]  
 RP SEQUENCE OF 331-382 FROM N.A.  
 RX MEDLINE=92407963; PubMed=1326639;  
 RA Chan S.Y., Bernard H.U., Ong C.K., Chan S.P., Birgit H., Delius H.;  
 RT "Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and  
 RT variants: a showcase for the molecular evolution of DNA viruses";  
 RL J. Virol. 66:5714-5725(1992).  
 CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF  
 CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2  
 CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH  
 CC CONTAINS BINDING SITES FOR BOTH PROTEINS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X74471; CAA52526.1; -  
 DR EMBL; M96321; AAA47010.1; -

DR PIR: S36493; S36493.  
 DR InterPro: IPR001177; -.  
 DR Pfam: PF00519; E1; 1.  
 DR Pfam: PF00524; E1.N; 1.  
 KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;  
 KW Nuclear protein.  
 FT NP\_BIND 432 439 ATP (POTENTIAL).  
 SQ SEQUENCE 604 AA; 68925 MW; 0A98B84E93B19B9 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 604;  
 Best Local Similarity 62.5%; Pred. No. 56;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 LHDDLEA 9  
 DB 201 VHDDLIES 208

RESULT 11  
 YDCP\_ECOLI  
 ID YDCP\_ECOLI STANDARD; PRT; 653 AA.  
 AC F76104; P76867; P76868; P76865;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DE SERINE/THREONINE-PROTEIN KINASE RIM15 (EC 2.7.1.-).  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PUTATIVE PROTEASE YDCP PRECURSOR (EC 3.4.-.-).  
 GN YDCP.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R. Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,  
 RA Oshima T., Saito N., Sampaio G., Seki Y., Sivasubram S.,  
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
 RA Yamamoto Y., Horiuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY U32.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE000240; AAC74517.1; AUT\_INIT.  
 DR EMBL: D90782; BAA15063.1; -.  
 DR EMBL: D90783; BAA15068.1; -.  
 DR EMBL: D90784; BAA15074.1; -.  
 DR EcoGene: EG13759; ydcP.  
 DR InterPro: IPR001539; -.  
 DR Pfam: PF01136; Peptidase\_U32; 1.

DR PROSITE: PS01276; PEPTIDASE\_U32; 1.  
 KW Hypothetical protein; Hydrolase; Protease; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 653 PUTATIVE PROTEASE YDCP.  
 SQ SEQUENCE 653 AA; 72701 MW; 5875B4E9C2F4FC82 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 653;  
 Best Local Similarity 75.0%; Pred. No. 61;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLHDDLE 8  
 DB 72 ILHDDLE 79

RESULT 12  
 RII5\_YEAST  
 ID RII5\_YEAST STANDARD; PRT; 1770 AA.  
 AC P43565;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SERINE/THREONINE-PROTEIN KINASE RIM15 (EC 2.7.1.-).  
 GN RIM15 OR TAK1 OR YFL033C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae.";  
 RL Nac. Genet. 10:261-268(1995).  
 [2]  
 SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC MEDLINE=97265402; PubMed=9111339;  
 RA Vidan S., Mitchell A.P.;  
 RT "Stimulation of yeast meiotic gene expression by the  
 RT glucose-repressible protein kinase Rim15p.";  
 RL Mol. Cell. Biol. 17:2688-2697(1997).  
 [3]  
 SEQUENCE FROM N.A.  
 RA Reinders A., Buerkert N., Boller T., Wienken A., de Virgilio C.;  
 RT "Saccharomyces cerevisiae CAMP-dependent protein kinase controls entry  
 RT into stationary phase through the Rim15p protein kinase.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: INVOLVED IN THE ACTIVATION OF A MEIOTIC GENES ACTIVATION  
 CC PATHWAY.

CC -1- PTM: AUTOPHOSPHORYLATED.  
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
 CC PROTEIN KINASES. STRONG, WITH S.POMBE CEK1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: D50617; BAA09206.1; -.  
 DR EMBL: U83459; AAB64088.1; -.  
 DR EMBL: AJ001030; CAA04486.1; -.  
 DR HSSP: Q63450; 1A06.  
 DR SGD: S0001861; RIM15.  
 DR InterPro: IPR000719; -.  
 DR InterPro: IPR001789; -.

DR InterPro: IPR002290; .  
 DR Pfam: PF00069; kinase; 2.  
 DR Pfam: PF00072; response\_reg; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR Transferrase: Serine/threonine-protein kinase; ATP-binding;  
 KW Phosphorylation; Meiosis.  
 FT DOMAIN 343 358 POLY-ASN.  
 FT DOMAIN 620 624 POLY-SER.  
 FT NP\_BIND 800 808 ATP (BY SIMILARITY).  
 FT BINDING 823 823 ATP (BY SIMILARITY).  
 FT ACT\_SITE 918 918 BY SIMILARITY.  
 FT DOMAIN 975 980 POLY-ASN.  
 FT DOMAIN 1213 1218 POLY-GLU.  
 FT DOMAIN 1386 1391 POLY-THR.  
 SQ SEQUENCE 1770 AA; 196530 MW; DC106482500FAFF CRC64;

Query Match 73.3%; Score 33; DB 1; Length 1770;  
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLHDDLE 8

Db 272 LLHNDLLE 279

RESULT 13  
 ID RLAL\_CHLRE STANDARD; PRT; 107 AA.  
 AC P29763;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 60S ACIDIC RIBOSOMAL PROTEIN P1.  
 OS Chlamydomonas reinhardtii.  
 CC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 CC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-137C;  
 RA Dumont F.;  
 RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF  
 CC -!- PROTEIN SYNTHESIS (BY SIMILARITY).  
 CC -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL  
 CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X66411; CAA47042.1; .  
 DR PIR: S24990; R6KMIC;  
 DR InterPro: IPR001813; .  
 DR Pfam: PF00428; 60s-ribosomal; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 107 AA; 10875 MW; 2A7FC9696D8617EF CRC64;

Query Match 71.1%; Score 32; DB 1; Length 107;  
 Best Local Similarity 75.0%; Pred. No. 12;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLHDDLE 8

Db 14 ILHDDGLE 21

RESULT 14  
 ID YVAE\_VACC STANDARD; PRT; 166 AA.  
 AC P20514;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE HYPOTHETICAL 18.2 KDA PROTEIN.  
 GN A ORF E.  
 OS Vaccinia virus (strain Copenhagen).  
 CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 CC Orthopoxvirus.  
 CC NCBI\_TaxID=10249;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021027; PubMed=2219722;  
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
 RA Paoletti E.;  
 RT "The complete DNA sequence of vaccinia virus.";  
 RL Virology 179:247-266(1990).  
 RN [2]  
 RP COMPLETE GENOME.  
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
 RA Paoletti E.;  
 RL Virology 179:517-563(1990).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M35027; AAA48130.1; .  
 DR PIR: A42524; A42524.  
 KW Hypothetical protein.  
 SQ SEQUENCE 166 AA; 18230 MW; 5927A96A0CB9477B CRC64;

Query Match 71.1%; Score 32; DB 1; Length 166;  
 Best Local Similarity 55.6%; Pred. No. 20;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLHDDLEA 9

Db 12 ILHDDCLKS 20

RESULT 15  
 ID BZLF\_EBV STANDARD; PRT; 245 AA.  
 AC P03206;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE BZLF1 TRANS-ACTIVATOR PROTEIN (EB1) (ZEBRA).  
 GN BZLF1.  
 OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).  
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Gammaherpesvirinae; Lymphocryptovirus.  
 CC NCBI\_TaxID=10377;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Sachwell S.C., Seguin C.,  
 RA Tuffnell P.S., Barrett B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";

```

RL Nature 310:207-211(1984).
RN [2]
RP IDENTIFICATION OF PROTEIN.
RX MEDLINE-87311873; PubMed-3041034;
RA Biggin M., Rodescot M., Ferricaudet M., Farrel P.;
RT "Epstein-Barr virus gene expression in P3HR1-superinfected Raji
RL cells.";
RL J. Virol. 61:3120-3132(1987).
RN [3]
RP DNA-BINDING.
RX MEDLINE-89231610; PubMed-2540954;
RA Farrel P.J., Rowe D.T., Rooney C.M., Kouzarides T.;
RT "Epstein-Barr virus BzLF1 trans-activator specifically binds to a
RL consensus AP-1 site and is related to c-fos.";
RL EMBO J. 8:127-132(1989).
RN [4]
RP STRUCTURE, AND FUNCTION.
RX MEDLINE-90219210; PubMed-2157874;
RA Packham G., Economou A., Rooney C.M., Rowe D.T., Farrel P.J.;
RT "Structure and function of the Epstein-Barr virus BZLF1 protein.";
RL J. Virol. 64:2110-2116(1990).
RN [5]
RP FUNCTION.
RX MEDLINE-94008999; PubMed-8404860;
RA Schepers A., Pich D., Hamerschmidt W.;
RT "A transcription factor with homology to the AP-1 family links RNA
RL transcription and DNA replication in the lytic cycle of Epstein-Barr
RL virus.";
RL EMBO J. 12:3921-3929(1993).
CC -1- FUNCTION: BZLF1 PLAYS A KEY ROLE IN THE SWITCH FROM LATENT TO
CC PRODUCTIVE INFECTION OF EBV. BZLF1 ACTIVATES THE PROMOTER OF AN
CC OTHER EBV GENE (BSLF2+BMFL1).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; CAA24861.1; -
DR PIR: A03768; QBE27.
DR PIR: S03634; S03634.
DR TRANSFAC: T00923; -
DR InterPro: IPR001871; -
DR Pfam: PF00170; bZIP; 1.
DR PROSITE: PS00036; BZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Early protein.
FT DNA_BIND 178 195 BASIC MOTIF.
SQ SEQUENCE 245 AA; 26860 MW; 7F3D55D79F1F0196 CRC64;

```

```

Query Match 71.1%; Score 32; DB 1; Length 245;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLHDDLL 7
Db 237 VLHEDLL 243

```

Search completed: July 10, 2001, 08:35:27  
Job time: 250 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:31:47 ; Search time 33.27 Seconds  
(without alignments)  
35.790 Million cell updates/sec

Title: US-09-489-760-2  
Perfect score: 45  
Sequence: 1 VLHDDLLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL16.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	84.4	1148	3 Q03897	Q03897 saccharomyc
2	37	82.2	1165	4 Q92619	Q92619 homo sapien
3	35	77.8	119	9 Q80057	Q80057 bacterioph
4	35	77.8	496	9 Q9MBP0	Q9MBP0 staphylococ
5	35	77.8	616	1 Q27025	Q27025 methanobact
6	34	75.6	332	5 Q9V966	Q9V966 drosophila
7	34	75.6	640	5 Q9NKM2	Q9NKM2 leishmania
8	34	75.6	741	5 Q25257	Q25257 leishmania
9	34	75.6	741	5 Q9NDK7	Q9NDK7 leishmania
10	34	75.6	774	5 Q9VA13	Q9VA13 drosophila
11	34	75.6	1383	5 Q9VDA0	Q9VDA0 drosophila
12	33	73.3	145	2 Q93387	Q93387 staphylococ
13	33	73.3	174	5 Q94256	Q94256 caenorhabdi
14	33	73.3	175	2 Q9RPT9	Q9RPT9 streptomyce
15	33	73.3	210	2 Q9S3K5	Q9S3K5 staphylococ
16	33	73.3	238	2 P74656	P74656 synechocyst
17	33	73.3	247	2 Q9SOL4	Q9SOL4 staphylococ
18	33	73.3	259	2 Q9PBQ0	Q9PBQ0 xylella fas
19	33	73.3	291	2 Q911R8	Q911R8 pseudomonas

20	33	73.3	319	14 Q9WHJ4	Q9WHJ4 walleye epi
21	33	73.3	320	2 Q9RVU0	Q9RVU0 deinococcus
22	33	73.3	322	2 Q9HVL5	Q9HVL5 pseudomonas
23	33	73.3	352	2 Q9RJH5	Q9RJH5 streptomyce
24	33	73.3	378	2 Q9X7V8	Q9X7V8 streptomyce
25	33	73.3	398	1 Q9YE58	Q9YE58 aeropyrum p
26	33	73.3	408	5 Q9V8W8	Q9V8W8 drosophila
27	33	73.3	416	10 Q9FWD0	Q9FWD0 oryza sativ
28	33	73.3	421	1 P94942	P94942 methanopyru
29	33	73.3	425	6 Q9GM45	Q9GM45 macaca fasc
30	33	73.3	654	2 Q9LAM6	Q9LAM6 salmonella
31	33	73.3	723	10 Q9SV54	Q9SV54 arabidopsis
32	33	73.3	856	10 Q9SV11	Q9SV11 arabidopsis
33	33	73.3	908	2 Q87123	Q87123 vibrio algi
34	33	73.3	973	3 Q9UR24	Q9UR24 schizosacch
35	33	73.3	1534	4 Q9UF04	Q9UF04 homo sapien
36	33	73.3	4150	2 Q9KIV4	Q9KIV4 streptomyce
37	32	71.1	91	2 Q07938	Q07938 bacillus su
38	32	71.1	166	14 Q9JF89	Q9JF89 vaccinia vi
39	32	71.1	240	11 Q63432	Q63432 rattus norv
40	32	71.1	245	14 Q69127	Q69127 human herpe
41	32	71.1	253	5 Q9N3G0	Q9N3G0 caenorhabdi
42	32	71.1	270	1 Q9H8H9	Q9H8H9 halobacteri
43	32	71.1	276	5 Q44591	Q44591 caenorhabdi
44	32	71.1	282	10 Q9FNI3	Q9FNI3 arabidopsis
45	32	71.1	294	1 Q54288	Q54288 sulfolobus

## ALIGNMENTS

## RESULT 1

Q03897 ID Q03897 PRELIMINARY; PRT; 1148 AA.  
AC Q03897;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HYPOTHEICAL 130.9 KDA PROTEIN.  
GN YDRL28W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID:4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB972;  
RA Oliver K., Harris D.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB972;  
RA Barrell B., Rajandream M.A.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z48179; CA88209.1; -  
DR SGB; S0002535; YDRL28W.  
DR InterPro; IPR001680; -  
DR Pfam; PF00400; WD40; 6.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR PROSITE; PS00678; WD\_REPEATS; UNKNOWN\_1.  
DR SMART; SM00320; WD40; 1.  
KW Repeat; WD repeat.  
SQ SEQUENCE 1148 AA; 130945 MW; D6A137FB0FDE2816 CRC64;

Query Match 84.4%; Score 38; DB 3; Length 1148;  
Best Local Similarity 66.7%; Pred. No. 86;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLHDDLLEA 9

Db 1001 LLHDDIIIEA 1009

```

RT "Panton-valentine leukocidin genes in a phage-like particle isolated
RT from mitomycin C-treated Staphylococcus aureus V8 (ATCC 49775).";
RL Biosci. Biotechnol. Biochem. 61:1960-1962(1997).
SQ SEQUENCE 119 AA; 13921 MW; 0F2A472DB7D63FA0 CRC64;

Query Match 77.8%; Score 35; DB 9; Length 119;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HDDLLEA 9
Db 72 HDDLLEA 78

RESULT 4
Q9MBP0 PRELIMINARY; PRT; 496 AA.
ID Q9MBP0;
AC Q9MBP0;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE PHI PVL ORF 18 AND 19 HOMOLOGUE.
OS Staphylococcus aureus prophage phiP83.
OC Viruses.
OX NCBI_TaxID=129009;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P83;
RA Zou D., Kaneko J., Narita S., Kamio Y.;
RT "Complete nucleotide sequence and molecular characterization of
RT prophage pV83pro carrying lukM-lukP-PV(p83) gene cluster in
RT Staphylococcus aureus strain P83.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P83;
RX MEDLINE=98067870; PubMed=9404084;
RA Kaneko J., Muramoto K., Kamio Y.;
RT "Gene of LukP-pv-like component of Panton-Valentine leukocidin in
RT Staphylococcus aureus p83 is linked with lukM.";
RL Biosci. Biotechnol. Biochem. 61:1960-1962(1997).
DR EMBL; AB044554; BAA97859.1; -.
SQ SEQUENCE 496 AA; 56878 MW; 1BE4F6E6D7920584 CRC64;

Query Match 77.8%; Score 35; DB 9; Length 496;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HDDLLEA 9
Db 72 HDDLLEA 78

RESULT 5
Q27025 PRELIMINARY; PRT; 616 AA.
ID Q27025
AC Q27025;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE TYPE I RESTRICTION MODIFICATION ENZYME, SUBUNIT M.
GN MTH942.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;

RT "Panton-valentine leukocidin genes in a phage-like particle isolated
RT from mitomycin C-treated Staphylococcus aureus V8 (ATCC 49775).";
RL Biosci. Biotechnol. Biochem. 61:1960-1962(1997).
SQ SEQUENCE 119 AA; 13921 MW; 0F2A472DB7D63FA0 CRC64;

Query Match 82.2%; Score 37; DB 4; Length 1165;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLHDDLLEA 9
Db 166 VLHDDLLEA 174

RESULT 3
O80057 PRELIMINARY; PRT; 119 AA.
ID O80057;
AC O80057;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE ORF 18.
OS bacteriophage phi PVL.
OC Viruses.
OX NCBI_TaxID=71366;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98067870; PubMed=9404084;
RA Kaneko J., Kimura T., Kawakami Y., Tomita T., Kamio Y.;

```



GenCore Version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:35:27 ; Search time 12.64 Seconds  
(without alignments)  
24.391 Million cell updates/sec

Title: US-09-489-760-5  
Perfect score: 42  
Sequence: 1 VLKDDLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	34	81.0	282	1 FU34_YEAST	P32907 saccharomyc
2	34	81.0	343	1 G3P_METJA	Q58546 methanococ
3	34	81.0	3519	1 RL56_STRAT	Q07017 streptomyce
4	33	78.6	465	1 RBL_DAPSP	P28397 daphniphyll
5	33	78.6	822	1 EPS8_HUMAN	Q12929 homo sapien
6	33	78.6	1041	1 SYI_MYCTU	Q10765 mycobacteri
7	33	78.6	1059	1 SYI_MYCLE	Q9x7e5 mycobacteri
8	32	76.2	264	1 YCV5_SCHPO	Q13679 schizosacch
9	32	76.2	334	1 NA37_ECOLI	P33920 escherichia
10	32	76.2	372	1 YC60_MYCTU	Q11058 mycobacteri
11	32	76.2	447	1 RBL_CONTR	Q05988 convolvulus
12	32	76.2	465	1 RBL_PLAOC	P28441 platanus oc
13	32	76.2	465	1 RBL_PROAR	P28458 trochodendr
14	32	76.2	499	1 AMPA_CHLPN	Q928f8 chlamydia p
15	32	76.2	598	1 CYSJ_ECOLI	P38038 escherichia
16	32	76.2	598	1 CYSJ_SALTY	P38039 salmonella
17	32	76.2	889	1 NIA3_MAIZE	P49102 zea mays (m
18	31	73.8	112	1 GUNB_RHOSH	P43519 rhodobacter
19	31	73.8	252	1 VMT1_IABAN	P03487 influenza a
20	31	73.8	252	1 VMT1_IACKB	P36347 influenza a
21	31	73.8	252	1 VMT1_IAPFW	P10918 influenza a
22	31	73.8	252	1 VMT1_IAPFR	P03488 influenza a
23	31	73.8	252	1 VMT1_IAPFW	P05775 influenza a
24	31	73.8	252	1 VMT1_IAMAN	P08381 influenza a
25	31	73.8	252	1 VMT1_IAUDO	P03486 influenza a
26	31	73.8	252	1 VMT1_IJAWL	P05777 influenza a
27	31	73.8	252	1 VMT1_IJAZ1	P05776 influenza a
28	31	73.8	441	1 RBL_FOUSP	Q05990 fouquieria
29	31	73.8	508	1 Y020_HUMAN	Q15397 homo sapien
30	31	73.8	593	1 CDR1_SCHPO	P07334 schizosacch
31	31	73.8	629	1 HAPI_RAT	P54256 rattus norv
32	31	73.8	682	1 RECG_BACSU	Q34942 bacillus su
33	31	73.8	862	1 GYRA_CAMFE	P47235 campylobact

34	31	73.8	921	1 SVI_BACSU	Q45477 bacillus su
35	31	73.8	1128	1 BEM3_YEAST	P32873 saccharomyc
36	31	73.8	1300	1 DYNA_NEUCR	Q01397 neurospora
37	31	73.8	2504	1 PAS_HUMAN	P49327 homo sapien
38	31	73.8	3149	1 TEGU_EBV	P03186 Epstein-Bar
39	30	71.4	195	1 NIFQ_AZOVI	P11088 azotobacter
40	30	71.4	231	1 YHHW_ECOLI	P46852 escherichia
41	30	71.4	267	1 EUTT_ECOLI	P76554 escherichia
42	30	71.4	267	1 EUTT_SALTY	Q92fV4 salmonella
43	30	71.4	338	1 CVF_PHOLA	P95522 phormidium
44	30	71.4	368	1 SERC_NEIMA	Q34370 neisseria m
45	30	71.4	368	1 SERC_NEIMB	P57007 neisseria m

## ALIGNMENTS

RESULT	1
FU34_YEAST	
ID	FU34_YEAST STANDARD; PRT; 282 AA.
AC	P32907
DT	01-OCT-1993 (Rel. 27, Created)
DT	01-OCT-1993 (Rel. 27, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	FUN34 PROTEIN.
GN	FUN34 OR YNR002C OR N2029.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX	NCBI_TaxID=4932;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GRF88;
RX	MEDLINE=93016077; PubMed=1400451;
RA	Stettler S., Mariotte S., Riva M., Sentenac A., Thuriaux P.;
RT	"An essential and specific subunit of RNA polymerase III (C) is
RT	encoded by gene RPC34 in Saccharomyces cerevisiae.";
RL	J. Biol. Chem. 267:21390-21395(1992).
RL	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GRF88;
RX	MEDLINE=95028151; PubMed=7941739;
RA	Lalo D., Stettler S., Mariotte S., Gendreau E., Thuriaux P.;
RT	"Organization of the centromeric region of chromosome XIV in
RT	Saccharomyces cerevisiae.";
RL	Yeast 10:523-533(1994).
RL	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=S288C / FY1679;
RX	MEDLINE=95208356; PubMed=7900425;
RA	Verhasselt P., Aert R., Voet M., Voickaert G.;
RT	"Twelve open reading frames revealed in the 23.6 kb segment flanking
RT	the centromere on the Saccharomyces cerevisiae chromosome XIV right
RL	arm.";
RL	Yeast 10:1355-1361(1994).
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC	-!- SIMILARITY: BELONGS TO THE GPRL/FUN34/YAAH FAMILY.
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X63746; CAA45279.1; -
DR	EMBL; X77395; CAA54571.1; -
DR	EMBL; Z71617; CAA96278.1; -
DR	PIR; S31258; S31258.
DR	SGD; S0005285; FUN34.
DR	InterPro; IPR000791; -
DR	Pfam; PF01184; Gp1_Fun34_Yaah; 1.

DR PROSITE: PS01114; GPR1\_FUN34\_YAAH; 1.  
KW Transmembrane.  
FT TRANSMEM 87 107 POTENTIAL.  
FT TRANSMEM 120 140 POTENTIAL.  
FT TRANSMEM 151 171 POTENTIAL.  
FT TRANSMEM 185 205 POTENTIAL.  
FT TRANSMEM 208 228 POTENTIAL.  
FT TRANSMEM 239 259 POTENTIAL.  
SQ SEQUENCE 282 AA; 30701 MW; F5E20F9324CE8199 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 282;  
Best Local Similarity 87.5%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 2 LRDLLEA 9  
| | | | |  
DB 62 LRDLLEA 69

RESULT 2  
G3P\_METJA STANDARD; PRT; 343 AA.  
AC Q58546;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 38, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).  
GN GAP OR Mcl146.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OX NCBI\_TaxID=21190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Uterback T.R., Kelley K.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,  
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii";  
RL Science 273:1058-1073(1996).

CC -|- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -|- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -|- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -|- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U67557; AB99147.1; -;  
CC TIGR: M1146; -;  
CC InterPro: IPR000173; -;  
CC Pfam: PF00044; gpdh; 1.  
CC PROSITE: PS00071; GAPDH; 1.  
CC Glycolysis; Oxidoreductase; NAD; NADP.  
CC BINDING 144 144 GLYCERALDEHYDE 3-PHOSPHATE (BY  
CC SIMILARITY).  
CC SEQUENCE 343 AA; 38102 MW; 81EB5810A9C838C5 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 343;  
Best Local Similarity 77.8%; Pred. No. 15;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 VLRDDLEA 9  
| | | | |  
DB 238 VSRDDLEA 246

RESULT 3  
OL56\_STRAT STANDARD; PRT; 3519 AA.  
AC Q07017;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE OLEANDOMYCIN POLYKETIDE SYNTHASE, MODULES 5 AND 6.  
GN ORFB.  
OS Streptomyces antibioticus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1890;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150470; PubMed=8107583;  
RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;  
RT "Characterisation of a Streptomyces antibioticus gene encoding a type  
I polyketide synthase which has an unusual coding sequence.";  
RL Mol. Gen. Genet. 242:358-362(1994).  
CC -|- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN  
CC LACTONE RING.  
CC -|- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.  
CC -|- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L09654; AA19695.1; -;  
CC HSSP: P25715; 1MLA.  
CC InterPro: IPR000255; -;  
CC InterPro: IPR000794; -;  
CC InterPro: IPR001031; -;  
CC InterPro: IPR001227; -;  
CC Pfam: PF00698; Acyl\_transf; 2.  
CC Pfam: PF00975; Thioesterase; 1.  
CC Pfam: PF00109; ketoacyl-synt; 2.  
CC Pfam: PF00550; pp-binding; 2.  
CC PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.  
CC PROSITE: PS00606; B-KETOACYL-SYNTHASE; 2.  
CC PROSITE: PS00075; ACP\_DOMAIN; 2.  
CC Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;  
KW Phosphopantetheine; Multifunctional enzyme.  
FT DOMAIN 1 ? MODULE 5.  
FT DOMAIN 2 ? MODULE 6.  
FT DOMAIN 3 3519 BETA-KETOACYL SYNTHASE (KS).  
FT DOMAIN 4 501 ACYLTRANSFERASE (AT).  
FT DOMAIN 5 569 BETA-KETOACYL REDUCTASE (KR).  
FT DOMAIN 6 1200 BETA-KETOACYL REDUCTASE (KR).  
FT DOMAIN 7 1487 ACYL CARRIER (ACP).  
FT DOMAIN 8 1686 BETA-KETOACYL SYNTHASE (KS).  
FT DOMAIN 9 2220 ACYLTRANSFERASE (AT).  
FT DOMAIN 10 2856 BETA-KETOACYL REDUCTASE (KR).  
FT DOMAIN 11 3141 ACYL CARRIER (ACP).  
FT DOMAIN 12 3215 ACYL-ENZYME INTERMEDIATE.  
FT ACT\_SITE ? ? PHOSPHOPANTETHEINE (BY SIMILARITY).  
FT BINDING ? ? THIOESTERASE.  
FT DOMAIN 3270 ?  
FT ACT\_SITE 210 BETA-KETOACYL SYNTHASE.

FT ACT\_SITE 660 660 ACYL-ENZYME INTERMEDIATE.  
 FT NP\_BIND 1203 1249 NADP.  
 FT BINDING 1524 1524 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT ACT\_SITE 1859 1859 BETA-KETOACYL SYNTHASE.  
 FT NP\_BIND 2311 2311 ACYL-ENZYME INTERMEDIATE.  
 FT NP\_BIND 2859 2905 NADP.  
 FT BINDING 3178 3178 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAE61F86 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 3519;  
 Best Local Similarity 87.5%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLEA 9  
 DB 770 LRDELLEA 777

RESULT 4  
 RBL\_DAPSP STANDARD; PRT; 465 AA.  
 ID P28397;  
 AC 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
 DE LARGE SUBUNIT) (FRAGMENT).  
 GN RBCL.  
 OS Chloroplast.  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Saxifragales;  
 OC Daphniphyllaceae; daphniphyllum.  
 OX NCBI\_TaxID=4388;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92397008; PubMed=1523408;  
 RA Albert V.A., Williams S.E., Chase M.W.;  
 FT "Carnivorous plants: phylogeny and structural evolution.";  
 RL Science 257:1491-1495(1992);  
 CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =  
 CC 2 3-PHOSPHO-D-GLYCERATE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =  
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; L01901; AAA84198.2; -  
 CC HSSP; P00875; LRXO.  
 CC Mendel; 5117; DAPSP; rbcl; 1.  
 CC InterPro; IPR000685; -  
 CC Pfam; PF00016; RUBISCO\_large; 1.  
 CC PROSITE; PS00157; RUBISCO\_LARGE; 1.  
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
 FT NON\_TER 1 1  
 FT ACT\_SITE 191 191 BINDING OF CO(2) ACTIVATES THE ENZYME.

SQ SEQUENCE 465 AA; 51606 MW; 019A80B3B25957EF CRC64;

Query Match 78.6%; Score 33; DB 1; Length 465;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRODLE 8  
 DB 338 LLRODLE 345

RESULT 5  
 EPS8\_HUMAN STANDARD; PRT; 822 AA.  
 ID Q12929;  
 AC 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8.  
 GN EPS8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94366758; PubMed=8084614;  
 RA Wong W.T., Carlomagno F., Druck T., Barletta C., Croce C.M.,  
 RA Huebner K., Kraus M.H., di Fiore P.P.;  
 FT "Evolutionary conservation of the EPS8 gene and its mapping to human  
 FT chromosome 1q23-q24.".  
 RL Oncogene 9:3057-3061(1994).  
 CC -!- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT  
 CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES ANALYZED, INCLUDING  
 CC HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND  
 CC PANCREAS. EXPRESSED IN ALL EPITHELIAL AND FIBROBLASTIC LINES  
 CC EXAMINED AND IN SOME, BUT NOT ALL, HEMATOPOIETIC CELLS.  
 CC -!- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SPLIT PH DOMAIN.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; U12535; AAA62280.1; -  
 CC HSSP; Q08509; LAOJ.  
 CC MIM; 600206; -  
 CC InterPro; IPR001452; -  
 CC Pfam; PF00018; SH3; 1.  
 CC PROSITE; PS50002; SH3; 1.  
 CC SH3 domain; Phosphorylation.  
 CC DOMAIN 69 129 PH (FIRST PART).  
 CC FT DOMAIN 210 213 POLY-PRO.  
 CC FT DOMAIN 322 325 POLY-PRO.  
 CC FT DOMAIN 381 414 PH (SECOND PART).  
 CC FT DOMAIN 421 440 PRO-RICH.  
 CC FT DOMAIN 532 591 SH3.  
 CC FT DOMAIN 615 651 PRO-RICH.  
 CC FT DOMAIN 659 664 POLY-SER.  
 CC SEQUENCE 822 AA; 91881 MW; AC5EB1D28B784B3B CRC64;

Query Match 78.6%; Score 33; DB 1; Length 822;  
 Best Local Similarity 75.0%; Pred. No. 60;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDLDLE 8  
 Db 552 VLKDDILE 559

```

RESULT 6
SYL_MYCTU STANDARD; PRT: 1041 AA.
ID SYL_MYCTU
AC Q10765; 006181;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5) (ISOLEUCINE--TRNA LIGASE)
DE (ILERS).
GN ILES OR RV1536 OR MTCY48.29C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaita F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + TRNA (ILE) = AMP +
CC PYROPHOSPHATE + L-ISOLEUCYL-TRNA (ILE).
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z74020; CA98326.1; -
CC TubercuList; RV1536; -
CC InterPro: IPR001412; -
CC InterPro: IPR002300; -
CC Pfam: PF00133; trna-synt_1; 1.
CC PRINTS: PR00984; TRNASYNTHILE.
CC PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
CC Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc.
CC SITE 53 63 "HIGH" REGION.
CC SITE 619 623 "KMSKS" REGION.
CC BINDING 622 622 ATP (BY SIMILARITY).
CC SEQUENCE 1041 AA; 117339 MW; B5023822848E08C6 CRC64;

Query Match 78.6%; Score 33; DB 1; Length 1041;
Best Local Similarity 77.8%; Pred. No. 78;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDLDLEA 9
Db 710 VLKDDLS 718
```

```

RESULT 7
SYL_MYCLE STANDARD; PRT: 1059 AA.
ID SYL_MYCLE
AC Q9X7E5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5) (ISOLEUCINE--TRNA LIGASE)
DE (ILERS).
GN ILES OR MLCB458.10.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K.J., Harris D., James K.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + TRNA (ILE) = AMP +
CC PYROPHOSPHATE + L-ISOLEUCYL-TRNA (ILE).
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL049478; CAB39575.1; -
CC InterPro: IPR002300; -
CC InterPro: IPR002301; -
CC Pfam: PF00133; trna-synt_1; 1.
CC PRINTS: PR00984; TRNASYNTHILE.
CC PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
CC Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc.
CC SITE 59 69 "HIGH" REGION.
CC SITE 637 641 "KMSKS" REGION.
CC BINDING 640 640 ATP (BY SIMILARITY).
CC SEQUENCE 1059 AA; 119808 MW; 67FC7659E9399E39 CRC64;

Query Match 78.6%; Score 33; DB 1; Length 1059;
Best Local Similarity 77.8%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDLDLEA 9
Db 728 VLKDDLSQA 736

RESULT 8
YCV5_SCHPO STANDARD; PRT: 264 AA.
ID YCV5_SCHPO
AC O13679;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 30.7 KDA PROTEIN C737.05 IN CHROMOSOME III.
DE SPCC737.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
```



Db 102 LLRDDLVE 109  
:|||||:

## RESULT 11

RBL\_CONTR  
ID RBL\_CONTR STANDARD; PRT; 447 AA.  
AC Q05988;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE RUBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
DE LARGE SUBUNIT) (FRAGMENT).  
DE RBCL.  
GN RBCL.  
OS Convolvulus tricolor.  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Convolvulaceae; Convolvulus.  
OX NCBI\_TaxID=4124;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Olmstead R.G., Michaels H.J., Scott K.M., Palmer J.D.;  
RT "Monophyly of the Asteridae and identification of their major lineages  
inferred from DNA sequences of rbcL";  
RL Ann. Mo. Bot. Gard. 79:249-265(1992).  
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
ACTIVE SITE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =  
2 3-PHOSPHO-D-GLYCERATE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =  
3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L11693; AAA84131.1; -  
CC HSP; P00876; 3RUB.  
CC InterPro; IPR000685; -  
CC Pfam; PF00016; RUBISCO\_large; 1.  
CC PROSITE; PS00157; RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 167  
FT ACT\_SITE 167 167 BINDING OF CO(2) ACTIVATES THE ENZYME  
FT (BY SIMILARITY).  
FT SEQUENCE 447 AA; 49663 MW; 6E56284E2C2DF270 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 447;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDDLLE 8

Db 314 LLRDDLVE 321

## RESULT 12

RBL\_PLAOC

ID RBL\_PLAOC STANDARD; PRT; 465 AA.  
AC P28441;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE RUBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
DE LARGE SUBUNIT) (FRAGMENT).  
DE RBCL.  
GN RBCL.  
OS Platanus occidentalis (Sycamore).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Platanaceae; Platanus.  
OX NCBI\_TaxID=4403;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=92397008; PubMed=1523408;  
RA Albert V.A., Williams S.E., Chase M.W.;  
RT "Carnivorous plants: phylogeny and structural evolution.";  
RL Science 257:1491-1495(1992).  
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
ACTIVE SITE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =  
2 3-PHOSPHO-D-GLYCERATE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =  
3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L01943; AAA84567.2; -  
CC HSP; P00876; 3RUB.  
CC InterPro; IPR000685; -  
CC Pfam; PF00016; RUBISCO\_large; 1.  
CC PROSITE; PS00157; RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 1  
FT ACT\_SITE 191 191 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT SEQUENCE 465 AA; 51691 MW; E9DB5903CCF3BBAD CRC64;

Query Match 76.2%; Score 32; DB 1; Length 465;  
Best Local Similarity 75.0%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDDLLE 8

Db 338 LLRDDLVE 345

## RESULT 13

RBL\_TROAR

ID RBL\_TROAR STANDARD; PRT; 465 AA.

AC P28458;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE RUBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO

DE LARGE SUBUNIT) (FRAGMENT).

DE RBCL.

OS Trochodendron aralioides.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Trochodendraceae; Trochodendroideae;  
 OC Trochodendron.  
 OX NCBI\_TaxID=4407;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92397008; PubMed=1523408;  
 RA Albert V.A., Williams S.E., Chase M.W.;  
 RT "Carnivorous plants: phylogeny and structural evolution."; Science 257:1491-1495(1992).  
 RL  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L01958; AA84695.2; -  
 CC HSSP: P00876; 3RUB.  
 CC Mendel: 2879; TROAR:rbcl;1.  
 CC InterPro: IPR000685; -  
 CC Pfam: PF001016; RUBISCO\_LARGE; 1.  
 CC PROSITE: PS00157; RUBISCO\_LARGE; 1.  
 CC Photosynthesis; Carbon dioxide fixation; Photorespiration;  
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
 FT NON\_TER 1 191  
 FT ACT\_SITE 1 191 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 FT SEQUENCE 465 AA; 51691 MW; E9DRS903CF3BBAD CRC64;

Query Match 76.2%; Score 32; DB 1; Length 465;  
 Best Local Similarity 75.0%; Pred. No. 51;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDLDLE 8  
 :|||||:  
 Db 338 LLRDLLE 345

RESULT 14  
 ID AMPA\_CHLPN STANDARD; PRT; 499 AA.  
 AC Q9Z8F8; Q9ZQJ2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE CYTOSOL AMINOPEPTIDASE (EC 3.4.11.1) (LEUCINE AMINOPEPTIDASE)  
 DE (LAP) (LEUCYL AMINOPEPTIDASE).  
 GN PEPA OR CPN0385 OR CP0370.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;

RT "genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39."; Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR  
 CC TURNOVER OF INTRACELLULAR PROTEINS. CATALYZES THE REMOVAL OF  
 CC UNSUBSTITUTED AMINO-TERMINAL AMINO ACIDS FROM VARIOUS PEPTIDES (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL AMINO ACID, XAA-1-  
 CC XBB-, IN WHICH XAA IS PREFERABLY LEU, BUT MAY BE OTHER AMINO ACIDS  
 CC INCLUDING PRO ALTHOUGH NOT ARG OR LYS, AND XBB MAY BE PRO.  
 CC -1- COFACTOR: MANGANESE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE  
 CC CYTOSOL AMINOPEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001623; AAD18529.1; -  
 CC EMBL: AE002199; AAF38219.1; -  
 CC EMBL: AP002546; BAA98593.1; -  
 CC MEROPS: M17.003; -  
 CC TIGR: CP0370; -  
 CC HSSP: P00727; ILAN.  
 CC InterPro: IPR000819; -  
 CC Pfam: PF00883; Peptidase\_M17; 1.  
 CC PRINTS: PR00481; LAMNOPPTDASE.  
 CC PROSITE: PS00631; CYTOSOL\_AP; 1.  
 KW Hydrolase; Amino-peptidase; Manganese.  
 FT METAL 263 263 MANGANESE (2) (BY SIMILARITY).  
 FT METAL 268 268 MANGANESE (1 AND 2) (BY SIMILARITY).  
 FT METAL 286 286 MANGANESE (2) (BY SIMILARITY).  
 FT METAL 345 345 MANGANESE (1) (BY SIMILARITY).  
 FT METAL 347 347 MANGANESE (1 AND 2) (BY SIMILARITY).  
 FT ACT\_SITE 275 275 POTENTIAL.  
 FT ACT\_SITE 349 349 POTENTIAL.  
 FT SEQUENCE 499 AA; 54509 MW; B0F281B1DF4B9EC2 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 499;  
 Best Local Similarity 77.8%; Pred. No. 55;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDLDLE 9  
 :||:||||  
 Db 395 VLAEDLLEA 403

RESULT 15  
 ID CYSL\_ECOLI STANDARD; PRT; 598 AA.  
 AC P38038; P14782;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE SULFITE REDUCTASE [NADPH] FLAVOPROTEIN ALPHA-COMPONENT (EC 1.8.1.2)  
 DE (SIR-FP).  
 GN CYSD.  
 OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B;  
 RA MEDLINE=93380164; PubMed=2550423;  
 RX Ostrowski J., Barber M.J., Rueger D.C., Miller B.E., Siegel L.M.,  
 RA Kredich N.M.;  
 RT "Characterization of the flavoprotein moieties of NADPH-sulfite  
 reductase from *Salmonella typhimurium* and *Escherichia coli*.  
 RT Physicochemical and catalytic properties, amino acid sequence deduced  
 RT from DNA sequence of *cysJ*, and comparison with NADPH-cytochrome P-450  
 RT reductase.";  
 RL J. Biol. Chem. 264:15796-15808(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE OF 593-598 FROM N.A.  
 RC STRAIN-B;  
 RX MEDLINE=89359425; PubMed=2670946;  
 RA Ostrowski J., Wu J.-Y., Rueger D.C., Miller B.E., Siegel L.M.,  
 RA Kredich N.M.;  
 RT "Characterization of the *cysJH* regions of *Salmonella typhimurium* and  
 RT *Escherichia coli* B. DNA sequences of *cysJ* and *cysH* and a model for  
 RT the *siroheme-Fds4* active center of sulfite reductase hemoprotein  
 RT based on amino acid homology with spinach nitrite reductase.";  
 RL J. Biol. Chem. 264:15726-15737(1989).  
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE 6-ELECTRON REDUCTION OF  
 CC SULFITE TO SULFIDE. THIS IS ONE OF SEVERAL ACTIVITIES REQUIRED  
 CC FOR THE BIOSYNTHESIS OF L-CYSTEINE FROM SULFATE. THE FLAVO-  
 CC PROTEIN COMPONENT CATALYZES THE ELECTRON FLOW FROM NADPH ->  
 CC FAD -> FMN TO THE HEMOPROTEIN COMPONENT.  
 CC -!- CATALYTIC ACTIVITY: HYDROGEN SULFIDE + 3 NADP(+) + 3 H(2)O =  
 CC SULFITE + 3 NADPH.  
 CC -!- COPACITOR: THIS SUBUNIT IS A FLAVOPROTEIN THAT BINDS ONE FMN AND  
 CC ONE FAD PER CHAIN.  
 CC -!- SUBUNIT: ALPHA(8)-BETA(4). THE ALPHA COMPONENT IS A FLAVOPROTEIN,  
 CC THE BETA COMPONENT IS A HEMOPROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M23008; AAA23650.1; -;  
 DR EMBL; U29579; AAA69274.1; -;  
 DR EMBL; AE000360; AAC75806.1; -;  
 DR PIR; B34231; B34231.  
 DR HSSP; P00388; IAMO.  
 DR EcoGene; EG10191; *cysJ*.  
 DR InterPro; IPR001094; -;  
 DR InterPro; IPR001433; -;  
 DR InterPro; IPR001709; -;  
 DR InterPro; IPR003097; -;  
 DR Pfam; PF00667; FAD\_binding; 1.  
 DR Pfam; PF00175; oxidored\_fad; 1.  
 DR PRINTS; PR00369; FLAVODOXIN.  
 DR PRINTS; PR00371; PFNCR.  
 KW Oxidoreductase; Flavoprotein; NADP; FAD; FMN; Electron transport;  
 KW Cysteine biosynthesis.  
 FT INIT\_MET. 0

FT NP\_BIND 69 73 FMN (RIBITYLPHOSPHATE) (BY SIMILARITY).  
 FT NP\_BIND 149 180 FMN (PYRIMIDINE PART) (BY SIMILARITY).  
 FT NP\_BIND 235 287 FAD (PYROPHOSPHATE MOIETY)  
 FT (BY SIMILARITY).  
 FT NP\_BIND 471 598 NADP (PYRIMIDINE MOIETY) (BY SIMILARITY).  
 FT CONFLICT 155 155 T -> S (IN REF. 2).  
 FT CONFLICT 267 267 L -> M (IN REF. 2).  
 FT CONFLICT 507 507 E -> D (IN REF. 2).  
 SQ SEQUENCE 598 AA; 66148 MW; 0B6EE0E2E70F13E CRC64;  
 Query Match 76.2%; Score 32; DB 1; Length 598;  
 Best Local Similarity 87.5%; Pred No. 67;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LRDDLLEA 9  
 Db 81 LRDDLLEA 88  
 Search completed: July 10, 2001, 08:35:27  
 Job time: 250 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:32:06 ; Search time 18.74 Seconds  
(without alignments)  
9.675 Million cell updates/sec

Title: us-09-489-760-5  
Perfect score: 42  
Sequence: 1 VLRDDLLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	78.6	1045	1	US-08-452-083-2
2	31	73.8	251	5	PCT-US95-12357A-1
3	31	73.8	508	2	US-08-861-464-10
4	31	73.8	508	2	US-08-396-001-10
5	31	73.8	508	4	US-09-323-433A-10
6	31	73.8	593	3	US-08-556-419-22
7	31	73.8	629	3	US-08-556-419-23
8	31	73.8	2509	1	US-08-469-005A-10
9	30	71.4	483	2	US-08-770-544-8
10	30	71.4	821	1	US-07-935-311A-4
11	30	71.4	821	1	US-08-368-079-4
12	30	71.4	821	5	PCT-US93-07996-4
13	30	71.4	1159	2	US-08-956-242-13
14	30	71.4	1159	3	US-09-351-215-13
15	30	71.4	1159	4	US-09-226-012-2
16	30	71.4	1159	4	US-09-226-012-4
17	29	69.0	283	2	US-08-701-191A-41
18	29	69.0	790	1	US-08-363-560-2
19	29	69.0	1410	4	US-09-335-409-3
20	29	69.0	2165	1	US-08-514-975B-2
21	29	69.0	2165	5	PCT-US95-12507-2
22	29	69.0	2713	5	PCT-US96-01735-1
23	28	66.7	240	1	US-07-985-668A-3
24	28	66.7	240	3	US-08-950-433-3
25	28	66.7	240	3	US-09-186-287-3
26	28	66.7	245	4	US-08-944-483-69
27	28	66.7	275	2	US-09-016-366A-17

28 28 66.7 275 2 US-08-978-404B-12 Sequence 12, Appl  
29 28 66.7 649 2 US-08-871-266B-16 Sequence 16, Appl  
30 28 66.7 649 2 US-09-018-864A-16 Sequence 16, Appl  
31 28 66.7 649 3 US-08-871-267B-22 Sequence 22, Appl  
32 28 66.7 1832 4 US-09-335-409-4 Sequence 4, Appl  
33 28 66.7 7257 4 US-09-335-409-5 Sequence 5, Appl  
34 27 64.3 9 3 US-09-217-609A-11 Sequence 11, Appl  
35 27 64.3 9 4 US-08-873-235B-11 Sequence 11, Appl  
36 27 64.3 120 2 US-08-232-087A-8 Sequence 8, Appl  
37 27 64.3 122 2 US-08-232-087A-7 Sequence 7, Appl  
38 27 64.3 139 1 US-07-846-992-5 Sequence 5, Appl  
39 27 64.3 139 1 US-08-469-555-5 Sequence 5, Appl  
40 27 64.3 140 1 US-07-846-992-4 Sequence 4, Appl  
41 27 64.3 140 1 US-07-846-992-6 Sequence 6, Appl  
42 27 64.3 140 1 US-08-469-555-4 Sequence 4, Appl  
43 27 64.3 140 1 US-08-469-555-6 Sequence 6, Appl  
44 27 64.3 200 1 US-07-855-412B-1 Sequence 1, Appl  
45 27 64.3 200 2 US-08-308-887A-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-452-083-2  
; Sequence 2, Application US/08452083  
; Patent No. 5756327  
; GENERAL INFORMATION:  
; APPLICANT: Sassanfar, Mandana  
; APPLICANT: Schimmel, Paul R.  
; TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL ISOLEUCYL-TRNA  
; TITLE OF INVENTION: SYNTHETASE GENES, TESTER STRAINS AND ASSAYS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,083  
; FILING DATE: 26-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/305,765  
; FILING DATE: 13-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brook, David E.  
; REGISTRATION NUMBER: 22,592  
; REFERENCE/DOCKET NUMBER: CPI94-08B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1045 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-452-083-2

Query Match 78.6%; Score 33; DB 1; Length 1045;  
Best Local Similarity 77.8%; Pred. No. 82;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLRDDLLEA 9

Db 720 VLRLDLS 728  
||||| 1:

## RESULT 2

PCT-US95-12357A-1  
; Sequence 1, Application PC/TUS9512357A  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude Children's Research Hospital  
; TITLE OF INVENTION: Nucleic Acid Encoding Mutant Matrix  
; TITLE OF INVENTION: Proteins Useful for Attenuation or Enhancement of  
; TITLE OF INVENTION: Influenza A Virus, Vaccines and Methods of making and  
; TITLE OF INVENTION: Using Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: PCT/US95/12357A  
; FILING DATE: 29-SEPT-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/316,419  
; FILING DATE: 30-SEPT-1994  
; CLASSIFICATION:  
; APPLICATION NUMBER: 08/471,100  
; FILING DATE: 6-JUNE-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fox, Samuel L.  
; REGISTRATION NUMBER: 30,353  
; REFERENCE/DOCKET NUMBER: 0656.048PC01  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 251 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
PCT-US95-12357A-1

Query Match 73.8%; Score 31; DB 5; Length 251;  
Best Local Similarity 85.7%; Pred No. 42;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
|:|||||

Db 228 LKDDLLE 234

## RESULT 3

US-08-861-464-10  
; Sequence 10, Application US/08861464  
; Patent No. 5874210  
; GENERAL INFORMATION:  
; APPLICANT: Guarente, Leonard P.  
; APPLICANT: Austriaco Jr., Nicanor  
; APPLICANT: Kennedy, Brian  
; TITLE OF INVENTION: Genes Determining Cellular Senescence  
; TITLE OF INVENTION: in Yeast

; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/861,464  
; FILING DATE: 22-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/396,001  
; FILING DATE: 28-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/09351  
; FILING DATE: 15-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/107,408  
; FILING DATE: 16-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: MIT-6408A22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 781-861-6240  
; TELEFAX: 781-861-9540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 588 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-861-464-10

Query Match 73.8%; Score 31; DB 2; Length 508;  
Best Local Similarity 85.7%; Pred. No. 93;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLLE 8  
|:|||||

Db 62 LRDDLVE 68

## RESULT 4

US-08-396-001-10  
; Sequence 10, Application US/08396001  
; Patent No. 5919618  
; GENERAL INFORMATION:  
; APPLICANT: Guarente, Leonard P.  
; APPLICANT: Austriaco Jr., Nicanor  
; APPLICANT: Claus, James  
; APPLICANT: Cole, Francesca  
; APPLICANT: Kennedy, Brian  
; TITLE OF INVENTION: Genes Determining Cellular Senescence  
; TITLE OF INVENTION: Yeast  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396.001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-396-001-10

```

```

Query Match      73.8%; Score 31; DB 2; Length 508;
Best Local Similarity 85.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LRDDLLE 8
Db      62 LRDDLVE 68

```

```

RESULT 5
US-09-323-433A-10
; Sequence 10, Application US/09323433A
; Patent No. 6218512
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James J.
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN YEAST
; FILE REFERENCE: 0050.1491-003
; CURRENT APPLICATION NUMBER: US/09/323.433A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: US 08/396,001
; PRIOR FILING DATE: 1995-02-28
; PRIOR APPLICATION NUMBER: PCT/US94/09351
; PRIOR FILING DATE: 1994-08-15
; PRIOR APPLICATION NUMBER: US 08/107,408
; PRIOR FILING DATE: 1993-08-16
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-323-433A-10

```

```

Query Match      73.8%; Score 31; DB 4; Length 508;
Best Local Similarity 85.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LRDDLLE 8
Db      62 LRDDLVE 68

```

```

RESULT 6
US-08-556-419-22

```

```

; Sequence 22, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony
; APPLICANT: Worley, Paul
; APPLICANT: Snyder, Solomon
; TITLE OF INVENTION: Huntingtin-associated protein
; FILE REFERENCE: 01107.52271
; CURRENT APPLICATION NUMBER: US/08/556.419C
; CURRENT FILING DATE: 1995-11-09
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 599
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-08-556-419-22

```

```

Query Match      73.8%; Score 31; DB 3; Length 599;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LRDDLLE 8
Db      205 LRDDLQ 211

```

```

RESULT 7
US-08-556-419-23
; Sequence 23, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony
; APPLICANT: Worley, Paul
; APPLICANT: Snyder, Solomon
; TITLE OF INVENTION: Huntingtin-associated protein
; FILE REFERENCE: 01107.52271
; CURRENT APPLICATION NUMBER: US/08/556.419C
; CURRENT FILING DATE: 1995-11-09
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-08-556-419-23

```

```

Query Match      73.8%; Score 31; DB 3; Length 629;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LRDDLLE 8
Db      205 LRDDLQ 211

```

```

RESULT 8
US-08-469-005A-10
; Sequence 10, Application US/08469005A
; Patent No. 5665874
; GENERAL INFORMATION:
; APPLICANT: KUHAJDA, FRANCIS P.
; APPLICANT: PASTERNAK, GARY A.

```

```

; TITLE OF INVENTION: CANCER RELATED ANTIGEN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,005A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/188,426
; FILING DATE: 24-JAN-1994
; APPLICATION NUMBER: 08/096,908
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: 07/917,716
; FILING DATE: 24-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 062482-0113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:

```

INFORMATION FOR SEQ ID NO: 10:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2509 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-469-005A-10

```

```

Query Match 73.8%; Score 31; DB 1; Length 2509;
Best Local Similarity 87.5%; Pred. No. 5.6e-02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 VLRRDLLE 8
    |||||
DB 1972 VLRRDLLE 1979

```

```

; RESULT 9
; US-08-770-544-8
; Sequence 8, Application US/08770544
; Patent No. 5907085
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Ling, Kai-Shu
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,544
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60009008
; FILING DATE: 21-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-770-544-8

```

```

Query Match 71.4%; Score 30; DB 2; Length 483;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 RDDLEA 9
    |||:|
DB 124 RDDLEA 130

```

```

; RESULT 10
; US-07-935-311A-4
; Sequence 4, Application US/07935311A
; Patent No. 5378809
; GENERAL INFORMATION:
; APPLICANT: Di Fiore, Pier Paolo
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: Substrate of the Epidermal Growth
; TITLE OF INVENTION: Factor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,311A
; FILING DATE: 19920825
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH035.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 821 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-935-311A-4

Query Match 71.4%; Score 30; DB 1; Length 821;  
Best Local Similarity 62.5%; Pred. No. 2.5e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDLDLLE 8  
|::|::|  
Db 551 VMKDDVLE 558

## RESULT 11

US-08-368-079-4  
Sequence 4, Application US/08368079

Patent No. 5610018

GENERAL INFORMATION:

APPLICANT: Di Fiore, Pier Paolo

APPLICANT: Fazioli, Francesca

TITLE OF INVENTION: eps8, A Substrate for the Epidermal Growth Factor

TITLE OF INVENTION: Kinase, Antibodies Thereto, and Methods of Use Thereof

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear

STREET: 620 Newport Center Drive, 16th Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/368,079

FILING DATE: 03-JAN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/935,311

FILING DATE: 25-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned A.

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH035.001DV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 821 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-368-079-4

Query Match 71.4%; Score 30; DB 1; Length 821;  
Best Local Similarity 62.5%; Pred. No. 2.5e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDLDLLE 8  
|::|::|  
Db 551 VMKDDVLE 558

## RESULT 12

PCT-US93-07996-4

Sequence 4, Application PC/TUS9307996

GENERAL INFORMATION:

APPLICANT: The Government of the United States, as represented by the

APPLICANT: Secretary of Health and Human Services

TITLE OF INVENTION: Substrate of the Epidermal Growth Factor Kinase

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear

STREET: 620 Newport Center Drive, 16th Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/07996

FILING DATE: 19930825

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 821 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US93-07996-4

Query Match 71.4%; Score 30; DB 5; Length 821;  
Best Local Similarity 62.5%; Pred. No. 2.5e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLRDLDLLE 8  
|::|::|  
Db 551 VMKDDVLE 558

## RESULT 13

US-08-956-242-13

Sequence 13, Application US/08956242C

Patent No. 5986081

GENERAL INFORMATION:

APPLICANT: Ganetzky, Barry S.

APPLICANT: Titus, Steven A.

TITLE OF INVENTION: Polynucleotides Encoding Herg-3

FILE REFERENCE: 960296.94550

CURRENT APPLICATION NUMBER: US/08/956,242C

CURRENT FILING DATE: 1997-10-22

NUMBER OF SEQ ID NOS: 13

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 13

TYPE: PRT

LENGTH: 1159

ORGANISM: Homo sapien

US-08-956-242-13

Query Match 71.4%; Score 30; DB 2; Length 1159;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RDDLLE 8  
|::|::|  
Db 835 RDDLLE 840

## RESULT 14

US-09-351-215-13

Sequence 13, Application US/09351215

; Patent No. 6087488  
; GENERAL INFORMATION:  
; APPLICANT: Ganetzky, Barry S.  
; TITLE OF INVENTION: Polynucleotides Encoding Herg-3  
; FILE REFERENCE: 960296.94550  
; CURRENT APPLICATION NUMBER: US/09/351.215  
; CURRENT FILING DATE: 1999-07-12  
; EARLIER APPLICATION NUMBER: 08/956.242  
; EARLIER FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 1159  
; TYPE: PRT  
; ORGANISM: Homo sapien  
; US-09-351-215-13

Query Match 71.4%; Score 30; DB 3; Length 1159;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RDDLLE 8  
Db 835 RDDLLE 840

RESULT 15  
US-09-226-012-2  
; Sequence 2, Application US/09226012  
; Patent No. 6207383  
; GENERAL INFORMATION:  
; APPLICANT: Keating, Mark T.  
; APPLICANT: Splawski, Igor  
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT  
; FILE REFERENCE: 2323-136  
; CURRENT APPLICATION NUMBER: US/09/226.012  
; CURRENT FILING DATE: 1999-01-06  
; EARLIER APPLICATION NUMBER: 09/122.847  
; EARLIER FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1159  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-226-012-2

Query Match 71.4%; Score 30; DB 4; Length 1159;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RDDLLE 8  
Db 835 RDDLLE 840

Search completed: July 10, 2001, 08:32:07  
Job time: 150 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2001, 08:36:07 ; Search time 33.27 seconds  
(without alignments)  
35.790 Million cell updates/sec

Title: US-09-489-760-5  
Perfect score: 42  
Sequence: 1 VLRDLEA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mhc:\*  
7: sp\_mammal:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	1765	4 Q92619	Q92619 homo sapien
2	38	90.5	1150	2 Q9KIV4	Q9KIV4 streptomyc
3	36	85.7	174	5 Q94256	Q94256 caenorhabdi
4	35	83.3	616	1 Q27025	Q27025 mechanobact
5	34	81.0	238	14 Q9WHX7	Q9WHX7 influenza a
6	34	81.0	239	14 Q9WHY9	Q9WHY9 influenza a
7	34	81.0	239	14 Q9WHY8	Q9WHY8 influenza a
8	34	81.0	239	14 Q9WHY7	Q9WHY7 influenza a
9	34	81.0	239	14 Q9WHY6	Q9WHY6 influenza a
10	34	81.0	239	14 Q9WHY5	Q9WHY5 influenza a
11	34	81.0	239	14 Q9WHY4	Q9WHY4 influenza a
12	34	81.0	239	14 Q9WHY3	Q9WHY3 influenza a
13	34	81.0	239	14 Q9WHY1	Q9WHY1 influenza a
14	34	81.0	239	14 Q9WHX9	Q9WHX9 influenza a
15	34	81.0	250	14 Q9WHY2	Q9WHY2 influenza a
16	34	81.0	252	14 Q93188	Q93188 influenza a
17	34	81.0	389	5 Q9XTD7	Q9XTD7 caenorhabdi
18	34	81.0	486	2 Q9PPU0	Q9PPU0 ureaplasma
19	34	81.0	791	5 Q9NG78	Q9NG78 caenorhabdi

20	34	81.0	3816	2 Q9KIV3	Q9KIV3 streptomyc
21	33	78.6	300	2 Q9KPG3	Q9KPG3 bacillus ha
22	33	78.6	324	10 P93619	P93619 vigna ungu
23	33	78.6	377	8 O09424	O09424 chrysosplen
24	33	78.6	392	8 Q36681	Q36681 quintinia v
25	33	78.6	396	8 O09418	O09418 chrysosplen
26	33	78.6	396	8 O09416	O09416 chrysosplen
27	33	78.6	432	8 O78605	O78605 isotria ver
28	33	78.6	432	8 O09416	O09416 chrysosplen
29	33	78.6	437	8 O09422	O09422 chrysosplen
30	33	78.6	440	8 Q31148	Q31148 saxifraga p
31	33	78.6	441	8 Q31822	Q31822 antrophum
32	33	78.6	455	8 Q31907	Q31907 bolandra or
33	33	78.6	455	8 Q32207	Q32207 elmera race
34	33	78.6	465	8 Q32484	Q32484 jepsonia pa
35	33	78.6	465	8 Q32436	Q32436 sullivantia
36	33	78.6	465	8 Q33509	Q33509 hydrangea m
37	33	78.6	466	8 O20245	O20245 achlys trip
38	33	78.6	467	8 Q9TLN7	Q9TLN7 asteropyrum
39	33	78.6	468	8 Q9GDP7	Q9GDP7 quintinia v
40	33	78.6	475	8 Q9MTF0	Q9MTF0 tofieldia p
41	33	78.6	833	10 Q91V01	Q91V01 arabidopsis
42	32	76.2	237	2 O06978	O06978 bacillus su
43	32	76.2	274	5 Q9W157	Q9W157 trophobla
44	32	76.2	389	2 O86581	O86581 streptomyc
45	32	76.2	452	3 O43088	O43088 schizosacch

## ALIGNMENTS

RESULT 1  
Q92619 PRELIMINARY; PRT; 1165 AA.  
ID Q92619  
AC Q92619; (Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE MYELOBLAST KIAA0223 (FRAGMENT).  
GN KIAA0223.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-BONE MARROW;  
RX MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,  
Tanaka A., Kotani H., Miyajima N., Nomura N.;  
RA "Prediction of the coding sequences of unidentified human genes. VI.  
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
RT analysis of cDNA clones from cell line KG-1 and brain.";  
RL DNA Res. 3:321-329(1996).  
RN (2)  
RP SEQUENCE OF 170-1165 FROM N.A.  
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,  
Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,  
Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,  
Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,  
Bruce R., Thomas P., Quan R., Kronmiller B., Arellano A.,  
Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,  
Carrano A.V.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: D86976; BAA13212.1; --  
DR EMBL: AC004151; AAC03237.1; --  
DR HSSP: Q07960; IRGP  
DR InterPro: IPR000198;  
DR InterPro: IPR002219;  
DR Pfam: PF00130; DAG-PE-bind; 1.  
DR Pfam: PF00620; RhoGAP; 1.  
DR PROSITE: PS00479; DAG\_PE\_BIND\_DOM\_1; UNKNOWN\_1.  
DR





DR PROSITE: PS50081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR SMART: SM00324; RhoGAP; 1.  
FT NON\_TER 1  
SQ SEQUENCE 1165 AA; 127344 MW; 92EF768CAF458C9 CRC64;

Query Match 100.0%; Score 42; DB 4; Length 1165;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLKDDLEEA 9  
DB 166 VLKDDLEEA 174

RESULT 2  
Q9KIV4 PRELIMINARY; PRT; 4150 AA.

ID Q9KIV4  
AC Q9KIV4  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE 8.8A-DEOXYOLEANDOLIDE SYNTHASE 1.  
GN OLEAI.  
OS Streptomyces antibioticus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1890;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363406; PubMed=10908114;  
RA Shan S., Xue Q., Tang L., Carney J.R., Belach M., McDaniel R.;  
RT "Cloning, Characterization, and Heterologous Expression of a  
RT Polyketide Synthase and P-450 Oxidase Involved in the Biosynthesis of  
RT the Antibiotic Oleandomycin."  
RL J. Antibiot. 53:502-508(2000).  
DR EMBL: AF220951; AAF82408.1;  
DR InterPro: IPR00169;  
DR InterPro: IPR00255;  
DR InterPro: IPR00794;  
DR InterPro: IPR001227;  
DR InterPro: IPR002106;  
DR Pfam: PF00109; Ketoacyl-synt; 3.  
DR Pfam: PF00550; pp-binding; 3.  
DR Pfam: PF00698; Acyl-trans; 3.  
DR PROSITE: PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
DR PROSITE: PS00075; ACP\_DOMAIN; 3.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 2.  
DR PROSITE: PS00012; PHOSPHOPANTHETINE; 3.  
DR PROSITE: PS00639; THYL\_PROTEASE\_HIS; UNKNOWN\_1.  
KW Phosphopantetheine; transferase.  
SQ SEQUENCE 4150 AA; 435286 MW; 9383296C4C16647D CRC64;

Query Match 90.5%; Score 38; DB 2; Length 4150;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRDDLEEA 9  
DB 1781 LRDDLEEA 1788

RESULT 3  
Q94256 PRELIMINARY; PRT; 174 AA.

ID Q94256  
AC Q94256  
DT FEB-1997 (TRENBLrel. 02, Created)  
DT FEB-1997 (TRENBLrel. 02, Last sequence update)  
DE 2001 (TRENBLrel. 16, Last annotation update)  
GN C. elegans.  
OS  
OC  
OC  
OX  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delta: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL: AE000868; AAB5440.1;  
DR InterPro: IPR002052;  
DR InterPro: IPR002296;  
DR InterPro: IPR003356;  
DR Pfam: PF02384; N6\_Mtase; 1.  
DR PRINTS: PR00507; N12N6MTFRASE.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BRISTOL N2;  
RC MEDLINE=94150718; PubMed=7906398;  
RA Wilton R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopda A., Saunders D., Shownkeen R.,  
RA Smailson N., Smith A., Sonhammer E., Staden K., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BRISTOL N2;  
RC Geisel C., Bradshaw H.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U64849; AAC9055.1;  
DR InterPro: IPR000004;  
DR SMART: SM00118; SAPB; 1.  
SQ SEQUENCE 174 AA; 19210 MW; A6303CF6383BBED9 CRC64;

Query Match 85.7%; Score 36; DB 5; Length 174;  
Best Local Similarity 77.8%; Pred. No. 23;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLKDDLEEA 9  
DB 129 VLKDDLEEA 137

RESULT 4  
Q27025 PRELIMINARY; PRT; 616 AA.

ID Q27025  
AC Q27025  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE TYPE I RESTRICTION MODIFICATION ENZYME, SUBUNIT M.  
GN MTH942.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=DELTA H;  
RC MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delta: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL: AE000868; AAB5440.1;  
DR InterPro: IPR002052;  
DR InterPro: IPR002296;  
DR InterPro: IPR003356;  
DR Pfam: PF02384; N6\_Mtase; 1.  
DR PRINTS: PR00507; N12N6MTFRASE.

